UNIVERSITY OF MICHIGAN
1954
SOCIAL SCIENCE LIBRARY

# Journal of Experimental Psychology

ARTHUR W. MELTON, Editor
AIR FORCE PERSONNEL AND TRAINING RESEARCH CENTER
LACKLAND AIR FORCE BASE
SAN ANTONIO, TEXAS

CONSULTING EDITORS

Judson S. Brown, *State University of Iowa*
Cletus J. Burke, *Indiana University*
Robert M. Gagné, *AF Personnel and Training Research Center*
W. R. Garner, *The Johns Hopkins University*
Frank A. Geldard, *University of Virginia*
James J. Gibson, *Cornell University*
Clarence H. Graham, *Columbia University*
David A. Grant, *University of Wisconsin*
Lloyd G. Humphreys, *AF Personnel and Training Research Center*
Arthur L. Irion, *Tulane University*
Donald B. Lindsley, *University of California, Los Angeles*
Neal E. Miller, *Yale University*
Edwin B. Newman, *Harvard University*
Kenneth W. Spence, *State University of Iowa*
Benton J. Underwood, *Northwestern University*
Delos D. Wickens, *Ohio State University*

Lorraine Bouthilet, Managing Editor

## CONTENTS

On Stereoscopic Depth Perception: K. N. OGLE ..... 225
Visual and Auditory Closure: B. W. WHITE ..... 234
Discrimination Transfer along a Pitch Continuum: R. A. BAKER AND S. W. OSGOOD ..... 241
Conspicuity of Flashing Light Signals of Different Frequency and Duration: S. J. GERATHEWOHL 247
The Acquisition of Compound Concepts as a Function of Previous Training: H. H. KENDLER AND R. VINEBERG 252
Rate of Response during Operant Discrimination: M. H. SMITH, JR. AND W. J. HOY ..... 259
Performance in Selective Learning as a Function of Hunger: C. K. RAMOND ..... 265
Extinction by Omission of Food as a Function of Goal-Box Confinement: W. C. STANLEY AND M. I. ROWE 271
A Test of One Aspect of Contiguity Theory: A. D. DAVIS ..... 275
The Conditioned Evocation of Attitudes (Cognitive Conditioning?): G. RAZRAN ..... 278
Set or Perceptual Defense? J. T. FREEMAN ..... 283
The Effect of Stimulus Complexity on Discriminative Responses: L. W. GREGG ..... 289
Extra-Task Performance as a Measure of Learning a Primary Task: H. P. BAHRICK, M. NOBLE, AND P. M. FITTS 298
The Learning of Sequential Dependencies: W. F. BENNETT, P. M. FITTS, AND M. NOBLE .. 303

American Psychological Association

Vol. 48 No. 4 October 1954

# JOURNAL OF EXPERIMENTAL PSYCHOLOGY

The JOURNAL OF EXPERIMENTAL PSYCHOLOGY is published monthly, two volumes per year, by the American Psychological Association, Inc. The subscription rate per volume is $7.50, or $15.00 annually. Single copies are $1.50. Subscriptions, orders, and business communications should be addressed to the American Psychological Association, Inc., 1333 Sixteenth St. N.W., Washington 6, D. C.

This JOURNAL publishes original experimental investigations which are intended to contribute toward the development of psychology as an experimental science. Studies with normal human subjects are favored over studies involving abnormal or animal subjects. Studies in applied experimental psychology or engineering psychology may be accepted if they have broad implications for experimental and theoretical psychology. Normally, articles of a length exceeding 15 printed pages cannot be accepted. However, an integrated series of studies accomplished simultaneously (e.g., most doctoral dissertations) must be presented in a single article, rather than in a series of articles.

Articles are published in the order of their receipt, except in rare circumstances. Authors are supplied with 50 free offprints without covers. All of the cost of an author's alterations in galley proof is charged to the author. Priority in publication is given to articles whose authors assume the full cost of publication. In 1954 the publication cost is $20 per printed page. Authors of priority publications receive no free offprints.

Address all articles submitted for regular or priority publication to the editor, Arthur W. Melton, Air Force Personnel and Training Research Center, Lackland Air Force Base, San Antonio, Texas. All manuscripts should be submitted in duplicate. Original figures are prepared for publication; duplicate figures may be photographic or pencil-drawn copies. Authors are cautioned to retain a copy of the manuscript to guard against loss in the mail.

To save printing costs supplementary material to articles in this Journal are deposited with the American Documentation Institute. Copies may be obtained by ordering the Document Number given in the footnote reference to ADI. Orders should be sent to ADI Auxiliary Publications Project, Photoduplication Service, Library of Congress, Washington 25, D. C. Orders *must* be accompanied by checks or money orders payable to the Chief, Photoduplication Service, Library of Congress.

**Manuscripts submitted to the JOURNAL *must* adhere to the conventions concerning reference citations, preparation of tables and figures, manuscript format, etc. as described in the Publication Manual of the American Psychological Association (*Psychol. Bull.*, 1952, 49 [4, Suppl.], 389–449). In particular, the organization of the manuscript should be text, references, footnotes, tables (one to a page), figure titles (typed on separate page), and figures. All typed material *must* be double-spaced with ample margins. The term "subject," with its various forms, is abbreviated *S*, *Ss*, *S*'s, and *Ss*'; analogous abbreviations are used for "experimenter" (*E*) and "observer" (*O*). When in doubt about the practices of the JOURNAL, authors should examine a recent issue.**

Entered as second-class matter, February 6, 1937, at the post office at Lancaster, Pa., under the Act of March 3, 1879. Acceptance for mailing at the special rate of postage provided for in paragraph (d-2), Section 34.40, P. L. & R. of 1948, authorized September 11, 1947.

Send address changes to 1333 Sixteenth St. N.W., Washington 6, D. C. Address changes must reach the Subscription Office by the 25th of the month to take effect the following month. Undelivered copies resulting from address changes will not be replaced; subscribers should notify the post office that they will guarantee second-class forwarding postage. Other claims for undelivered copies must be made within four months of publication.

Copyright, 1954, by the American Psychological Association, Inc.

Journal of
Experimental Psychology

VOL. 48, No. 4 OCTOBER, 1954

# ON STEREOSCOPIC DEPTH PERCEPTION

KENNETH N. OGLE

*Section of Biophysics and Biophysical Research, Mayo Clinic and Mayo Foundation*[1]

The relationship between retinal image disparity and stereoscopic depth perception is well known. Moreover, there is now fairly general agreement on the nature of this relationship and the importance of disparity as the sole stimulus that gives rise to the stereoscopic experience. However, from time to time the present point of view is disputed. Just recently Tausch (4, p. 418) has done so with the conclusion that (in translation) ". . . the binocular perception of space does not result directly from disparity or correspondence of stimulated retinal elements, but rather because the stimuli received by each eye are transformed into forms (Gestalt). . . . The theory of local depth signs of disparately stimulated retinal elements cannot be upheld. . . . as far as optics are concerned, we are dealing with figure-formed processes (corresponding to the figure characteristics) and with figure-formed perception. . . ." These statements take one back to an early fundamental experiment of van der Meulen and van Dooremaal (5, p. 141), from which it was concluded that "stereoscopic depth is not to be conceived as a direct physiologic phenomenon, but as a psychic one, that is, the depth is produced by a psychophysical process." Now it is clear from reading their papers that in 1873 these authors implied in the use of the term "psychophysical" a perception based on an elaboration of stimuli from the two eyes as a central process, and that consequently stereoscopic depth was not a sensation directly derived from a physiological response to disparity as the stimulus.

Van der Meulen and van Dooremaal used the Hering "Fallversuch," or falling sphere test. The *O* viewed through a horizontal slit-aperture a portion of the thread of a plumb line and judged whether a small bead or sphere dropped by an assistant in the vicinity of the plumb line fell nearer or farther than the plumb line. This is a sensitive test for stereoscopic depth and it eliminates many of the empirical clues to depth perception. These *E*s placed a prism, edge up, before one eye, of such strength that the image of the slit-aperture seen by that eye was displaced entirely above the actual slit seen by the other eye. The *O* continued to use the plumb line

[1] The Mayo Foundation is a part of the Graduate School of the University of Minnesota.

as a fixation object.[2] When the sphere was now dropped near the plumb line the upper part of the path was seen by one eye, the lower part of the path by the other eye. In spite of the fact that at no instant was the sphere seen by both eyes simultaneously, the *E*s reported that the path of the falling sphere could nevertheless almost always be judged correctly in front of or behind the plumb line. They concluded that in this experiment "the stereoscopic depth experience arises solely through the means of *imagined prolongations* of the half-images, which then in the usual manner would be referred to horizontally corresponding (or disparate) retinal points" (**5**, p. 440).

Now Hofmann (1) apparently repeated the experiment, for he wrote that the depth impression from the falling sphere, while sufficiently convincing, was nevertheless not very obvious. He wrote further that if this depth sense is always the case "then we would have to classify binocular depth perception among the Gestalt perceptions. The central factor involved in this perception would be a behavior acquired by experience in the life of the individual, as unconscious, effectual reaction such as we have already learned to think of as the basis for Gestalt perception" (**1**, p. 440).

The experiment of van der Meulen and van Dooremaal, so far as the writer can learn, has never been studied under the controlled conditions possible with modern instrumentation. In this paper experiments will be described which extend the basic experiment of van der Meulen and van Dooremaal. The results confirm their experiment only to a very restricted degree, and on the whole show that for the emergence of a true stereoscopic depth experience the disparate images must fall on specifically horizontally associated retinal elements. This conclusion would throw doubt on a purely psychic basis for stereoscopic depth perception.

[2] It is not clear how this was done unless these *E*s had such a voluntary control of their convergence that they could keep the images of the plumb line aligned.

## Instrumentation

A schematic perspective drawing of the instrument is shown in Fig. 1. The *O*'s head was held in the instrument by a headrest and wax bite. The binocular visual field was restricted to a fixation point (a brightly illuminated spot on a sheet of plate glass 50 cm. from the eyes) and to a reference object and a test object seen by reflection from suitable 50–50 transmitting-reflecting mirrors. Polished thin steel drill rods, 4 cm. long, mounted vertically were used for the reference and test objects. These, when suitably illuminated, produced brilliant thin vertical lines. The reference rod was seen by reflection from the fixed mirror and could be adjusted so as to be seen by *O* at any desired position in front of or behind the plane of the fixation point and at any peripheral angle up to about 8°. The distance of the test object, which was seen by reflection from the double mirror arrangement, was kept constant. It could be adjusted for lateral angles only.

The disparity between the images of the test object, and therefore the stereoscopic depth as seen by *O*, could be changed by a small rotation of the two mirrors about vertical axes (A . . . A), through small angles in directions opposite to each other. The rotations of the two mirrors were controlled by suitable levers and a screw, which could be operated by a hand knob by *O* or by *E*. The angle of disparity between the images of the test line, for the small angles used, would be approximately proportional to the amount of rotation of the mirrors. This rotation could be read by a dial indicator. The dial was calibrated to read minutes of arc, and an estimated $\frac{1}{2}$ scale division corresponded to $\frac{1}{15}$ min. of arc. The dial indicator was set for zero when the test object appeared the same distance as the fixation point.

This means of presenting the test object made possible a wide range of image disparities. Most important was the fact that as the disparity was changed with the corresponding change in stereoscopic depth, the images did not change angular size, brightness, or sharpness because their optical distances did not change.

Thus empirical clues to spatial depth associated with the test object were essentially avoided.[3]

The reference and test objects were illuminated by lamps on the same electric circuit. They could be illuminated momentarily by a time switch in the current or could be continuously illuminated.

In order to repeat the van der Meulen and van Dooremaal experiment, a device for dropping a small sphere could be substituted for the test drill rod. This device consisted of a reservoir for holding about 100 polished steel ball bearings ($\frac{1}{16}$ in. diameter) and a cable release arrangement that caused a sphere to fall into a cup below the visual field of *O*.

The fixation spot, the reference and test lines were observed against a large, evenly illuminated white screen. The apertures that restricted the field of vision were illuminated to about the same brightness as the background. The actual backgrounds for the reference and test objects were mat black. The experiments otherwise were conducted in a dimly lighted room.

The important part of the instrument was a baffle device interposed in front of the test object in order to screen certain portions of the test rod from each of the two eyes. The baffles were two thin square pieces of blackened metal, with inner edges beveled. These were attached to a suitable mechanism that permitted any vertical or horizontal adjustment of each baffle plate. By means of scales and verniers, the vertical separation of the baffle plates could be adjusted accurately. In general, the baffle plates were adjusted so that one eye saw the entire upper portion of the test line, but not the lower, while the other eye saw the entire lower portion of the test line but not the upper portion. When there was a vertical gap, both eyes could see a center portion of the test object through it. The side aperture cut from black cardboard prevented the eyes from seeing the two ends of the test line. Thus each eye saw the same total length of the test line, including but independent of the height of the gap through which both eyes saw the central portion of the test line.

Care must be taken in the adjustment of the baffle plates for the determination of the zero gap. In this lies an important source of error. Because this would be influenced by the size of the pupils, care was taken not to change the general adaptation level of the eyes during the experiments and during the determination of the gap zero. The *O*, with head in the instrument, adjusted the lower baffle plate so that its inner edge appeared to coincide with the center of the fixation spot. With the test drill rod illuminated, or a sheet of white paper behind the rod illuminated, the upper baffle plate was then adjusted vertically until its inner edge just eliminated the light through the gap. The scales and verniers were then read and recorded. Each of the baffle plates was displaced one-half any desired gap width.

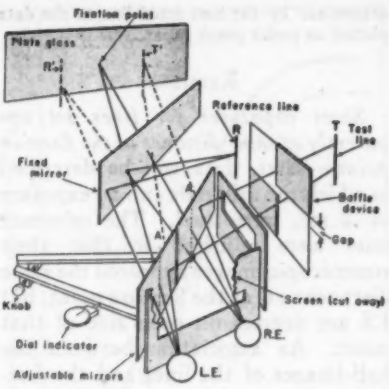


FIG. 1. Drawing of instrument in perspective. (Reproduced with permission from Ogle, K. N. Basis of stereoscopic vision. *A.M.A. Arch. Ophth.*, 1954, 52, 197–211.)

The principal purpose of the experiments was to determine the stereoscopic sensitivity of the eyes in the adjustment of the test line as the gap width was changed, that is, as the length of the binocularly seen portion of the test line was reduced. All the data were taken by two *O*s. One was an *O* experienced in stereoscopic tests. To correct a partial presbyopia and to reduce the accompanying esophoria, +1.25 diopter spheres were used before both eyes. The other *O* had had only a limited experience in stereoscopic observation. His eyes were myopic by .50 diopter, but no spectacle corrective lenses were used in these observations.

For the most part the method of adjustment was used. The stereoscopic sensitivity was taken as the reciprocal of the *SD* of the disparity in minutes of arc. For the second *O* considerable data were obtained using a forced-choice "near" or "far" constant stimulus method. The stereoscopic sensitivity was again taken as the reciprocal of the *SD*, which in turn was

[3] The very absence of empirical clues, of course, may itself constitute a negative empirical factor, which under certain circumstances might conceivably depress or dampen the effect of the disparity alone.

determined by the best fitted line to the data plotted on probit graph paper.

## Results

*Short exposures for lines set apparently at same distance as the fixation point.*—First, data will be described in which the drill rods and an exposure of .2 sec. were used. The reference lines were adjusted so that their stereoscopic images appeared the same distance to *O* as the fixation point, but 1.5 arc degrees on each side of that point. An association between the half-images of the lines and the fixation point was prevented by this lateral placement of the two lines. The timer in the circuit was adjusted for .2-sec. exposure of the two lines, when the button was pushed. This exposure was considered long enough to assure an adequate stimulus for stereoscopic perception, but at the same time short enough to prevent any appreciable influence of involuntary eye movements.

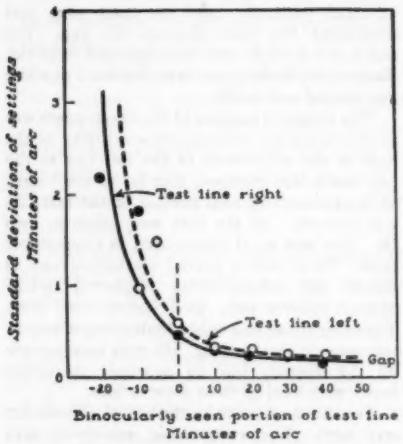


Fig. 2. The change in *SD* of settings as the binocularly seen portion of a test line is decreased. Settings made relative to reference line at same distance as fixation point ($\eta = 0$). Peripheral angle = 1°; reference and test lines exposed .2 sec.

With the method of adjustment, *O* adjusted the knob following exposures by a series of successive approximations until the test line appeared the same distance as did the fixation point and reference line. The dial indicator was read and recorded. The disparity was then thrown out of adjustment. This process was repeated ten times (at least), with short rest periods. The mean and *SD*'s in minutes of arc were computed from the data. The stereoscopic sensitivity was computed as the reciprocal of the *SD*.

The baffle plates were first adjusted for a gap of 4 mm., which corresponds to a vertical visual angle of the test line seen binocularly of 40 min. of arc. Data were obtained for a series of steps of decreased vertical gap, until stereoscopic depth could no longer be perceived. The positions of the reference and test lines were then interchanged, and the entire experiment was repeated.

From general observations it was clear that the settings were relatively easy to make as long as a portion of the test line could be seen binocularly. After the gap was closed, however, these settings became increasingly difficult, in that considerable changes in disparity could not be perceived as a difference in depth, though judgments of "nearer" or "farther" than the fixation point were more certain if the disparities were not too small. If the gap was further closed, so that the two half-images were actually separated, the settings became very difficult to make, and for larger separations even all sense of "nearer" or "farther" was lost. Thus under the conditions of the experiment, with short exposures of test object, the basic result of van der Meulen and van Dooremaal, when no portion of the test object is seen binocularly, was confirmed only within very narrow limits, and then the depth sense was indeed vague and indefinite, certainly more so than their paper suggests.

Quantitatively the typical results of one *O* are illustrated in Fig. 2. In this the angular sizes of the gaps or of the binocularly seen portions of the

test line are plotted as the abscissas, and the *SD*'s of the settings are plotted as the ordinates. The two sets, when the test line is seen on the left and when it is seen on the right side of the fixation point, are shown. Inspection of the figure shows that the variation in settings slowly increases as the gap closes, and then more precipitously after it closes, together with a considerable loss of consistency.

The difference between the two sets of data may be accounted for by (*a*) a possible cyclotorsional position of the eyes, (*b*) an asymmetry between retinal organization on two sides of the fixation point, but probably more by (*c*) an error in zero adjustment of gap, and (*d*) error in placing test line at the same lateral angle from fixation point.

If the reciprocals of the *SD*, which would correspond to the stereoscopic acuity or sensitivity, are plotted, Fig. 3 is obtained. Obviously, small errors in the *SD* when the gap is large will be exaggerated in the reciprocal plot. Similarly, the large errors when the gap is small or negative will be minimized. This graph shows that the stereoscopic sensitivity decreases almost linearly below a certain gap width. Caution should be exercised in assuming that this linearity continues to the abscissal axis. Nor is the fact that the straight line appears to continue after the gap is closed evidence that the stereoscopic processes on the two sides of the gap = 0 are the same.

Now the fact that there is a depth perception with an actual separation (within limits) of the half-images might be explained on the basis that there are vertical extents of areas for stereoscopic vision, analogous to the vertical extent of Panum's areas of fusion. That is, horizontally disparate images will be perceived in

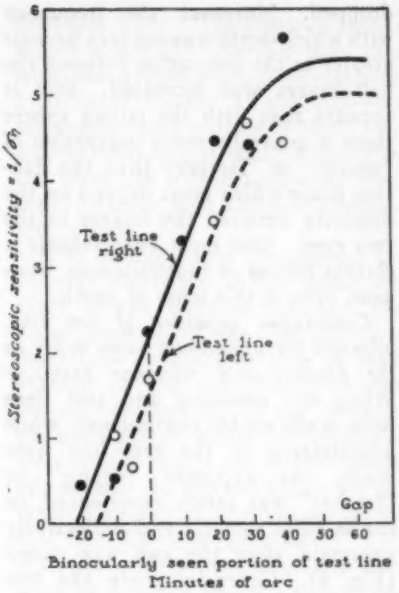
Stereoscopic sensitivity = 1/σ
Test line right
Test line left
Gap
-20 -10 0 10 20 30 40 50 60
Binocularly seen portion of test line
Minutes of arc

Fig. 3. Illustration of same data as those of Fig. 2, except that stereoscopic sensitivity, that is, the reciprocal of the *SD*, is plotted as ordinate

depth even when a small vertical disparity also exists between those images. Limiting horizontal extents of such areas of stereoscopic depth have already been shown to exist (3). From Fig. 3 one would estimate the vertical extent of such area for an excentric angle of 1.5° to be about 18 min. of arc. This angle seems offhand too large to be explained as an area effect, but judgment must be reserved on this point.

When the falling sphere device was substituted for the drill rod, the depth perception after the gap closed seemed more certain, and somewhat larger separations of the half-images could be used before all depth perception failed. However, even with the falling sphere, a depth of "nearer" or "farther" than the fixation point was not experienced with every sphere

dropped. Moreover the frequency with which depth was not seen became greater as the separation between the half-images was increased. Still it appears that with the falling sphere there is generally some conception of "nearer" or "farther" than the fixation plane which must depend on the disparity between the images in the two eyes. One must acknowledge a certain feeling of indefiniteness, however, even in this sense of depth.

*Continuous exposure of test line, adjusted for apparently same depth as the fixation and reference points.*—When the reference and test lines were made visible continuously while adjustments of the test line were made, the apparent "nearer" or "farther" was rarely experienced, or it was quite variable and subjectively uncertain after the gap was closed (Fig. 4). Generally when the test line was continuously visible, the stereoscopic sense of depth faded easily, or was prevented by the occurrence of suppression. As the angle of disparity was increased, these troubles seemed to increase. Blinking of the eyes was constantly resorted to to make the images reappear. There was the ever-present impulse to look at the reference and test lines. Little or no depth was experienced after the gap was closed.

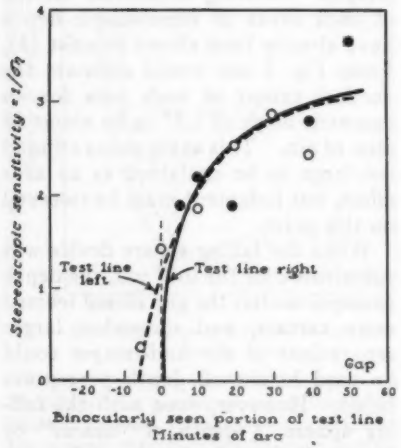


FIG. 4. Data from same experiment except that the reference and test lines were continuously visible

*Reference and test lines in front of (or behind) the plane of the fixation point; short exposures.*—An undesirable feature of the experiments just described is a lateral displacement between the upper and lower half-images when there is a disparity between those images. This displacement was an annoying factor that frequently interfered with the task of depth discrimination. The *O* learned that this displacement does not exist when the disparity is zero, and furthermore that at this zero disparity the test line should appear at the same distance as the fixation point. The falling sphere also appeared to jump laterally at the gap, when there is a disparity between the two images.

Final conclusions should not be drawn from the data or observations obtained when the reference line lies at the same distance as the fixation point. First, it is possible that the depth experienced when there is an actual separation between the two half-images may occur only for small horizontal disparities, that is, for objects near the frontoparallel plane of the fixation point, and would be less evident for objects at some distance in front of or behind that plane. The observations from the previous experiment might suggest that while the disparity from separated images may provide a depth experience of "nearer" or "farther," it may not result in an experience of a definite depth interval as always occurs with binocularly seen, vertically unseparated images.

Both to avoid the displacement factor and to test these possibilities, the reference line was placed a given distance in front of (or behind) the

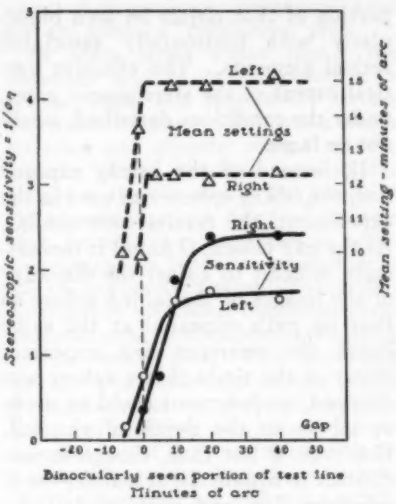


FIG. 5. The stereoscopic sensitivity and mean settings of reference and test lines seen in 13 min. of arc crossed disparity. Peripheral angle = 5°; exposure duration of lines = .2 sec.

frontoparallel plane of the fixation point. The *O* was then asked to adjust the disparity between the images of the test line so that the test line appeared at the same distance as the reference line, while *O* still maintained fixation on the original fixation spot. Then both the images of the reference and test lines would be approximately equally crossed (or uncrossed) disparate.

It became evident at once that even with the short exposure the test line could not in general be easily adjusted to appear at the same depth as the reference line, after the gap closed. Typical data showing this are illustrated in Fig. 5. For these particular data the reference and test lines were set at a peripheral angle of $\alpha_0 = 5°$ on either side of the fixation point. Also, they were set approximately 1.5 cm. in front of the fixation point, which corresponds to a crossed disparity of $\eta_0 = 13$ min. of arc. Again the duration of the exposure of the two lines was .2 sec. The maximal stereoscopic sensitivity was less than that when the reference disparity was zero. It dropped rapidly as the gap—the binocularly seen portion of test line—was narrowed, and became nonexistent after the half-images were separated by more than about 4 (average) min. of arc.

Also in this figure are plotted the angular disparities of the mean settings of the test-line images. The mean position of the test line was essentially constant as long as some portion of the test line was seen binocularly; but as the gap between the baffle plates closed, the validity of stereoscopic depth perception generally failed. In this particular set the mean disparity of the images of the test line suddenly decreased.[4] This behavior of the mean disparity data must be due to a loss of a patent and valid stereoscopic depth perception and the utilization of empirical factors. The depth of the test line from the separated half-images and that of the reference line may have been equated on the basis of apparent brightness, lengths, or some other factor associated with the half-images.

The decreased values of the disparity corresponding to the mean settings were usually obtained for the one *O* (KNO), though occasionally in a given set of data it was less evident. For the other *O* (RH), the change in mean disparity behavior was more likely to be in the opposite direction; that is, the mean disparity suddenly increased, in order for the test line to appear at the same depth as the reference line. Figure 6 illustrates

[4] The difference in the mean disparities curves (about 3 min. of arc) between test line right and left can be attributed to a difference in the magnification of the images in the two eyes (R < L) by about .5%. The difference in maximal stereoscopic sensitivity cannot be easily explained except that the two sets of data were obtained on different days.

data obtained by this *O*, for a peripheral angle of 1°, and a crossed disparity of the images of the reference line of 17.5 min. of arc. These data were obtained by the method of constant stimulus (forced "far" or "near"). The curve for the stereoscopic sensitivity suggests a stereoscopic depth after a separation between the half-images occurs. However, these data are very uncertain and probably represent a depth experience based on factors other than stereopsis, since the validity of the stereoscopic depth was so much in error.

This break in the mean disparities may be more important evidence of the loss of patent stereopsis than the decrease in the stereoscopic sensitivity. Obviously these data could not be obtained when the disparity between the images of the reference line was zero, i.e., the reference line was set in the frontoparallel plane of the fixation point.

These results provide evidence that a valid stereoscopic depth discrimination of an object requires that a portion of that object be seen binocularly with horizontally associated retinal elements. The effective vertical extent of the stereoscopic areas, under the conditions described, would not be large.

If, instead of the briefly exposed line, the falling sphere was used in the experiment, the results were similar. As the gap closed, *O* found it increasingly difficult to adjust the disparity of the images of the falling sphere so that its path appeared at the same depth the reference line appeared. Many of the times that a sphere was dropped, no judgment could be made at all as to the depth of the fall. However, *O* felt that, though it was difficult to equate the actual depths of reference line and path of fall, he nevertheless was aware that, within limits, the fall was "nearer" (or "farther") than the fixation point.

When the reference and test lines were continuously visible, the judgments of depth differences became very difficult as the gap was narrowed. Suppression of the half-images was especially noticeable.

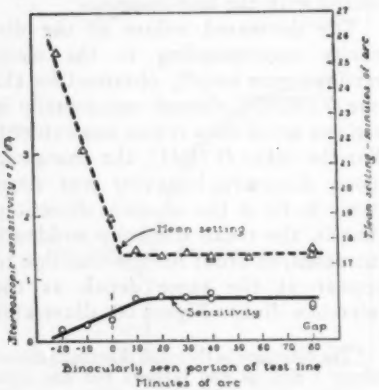


FIG. 6. The stereoscopic acuity and mean settings for the falling sphere and reference lines seen in 17.5 min. of arc crossed disparity. Peripheral angle = 1°; reference line on right side; constant stimulus method used.

## DISCUSSION

Although only two *O*s were used in the experiments just described, sufficient data were obtained and observations made to warrant some conclusions regarding stereoscopic vision, at least under the conditions imposed by the experimental procedure used.

The results clearly suggest that there may be two separate aspects of binocular depth perception, namely, a patent quantitative stereoscopic sense of depth and a qualitative depth experience, both of which, however, depend on the disparity between the images of the two eyes. These two aspects were probably identical to the "patent" stereopsis and qualitative experience of depth already suggested by the experiments on the disparity limits of stereopsis (3). It would be preferable to call the first

aspect of stereopsis an "obligatory" type and the second a "facultative" type.

As these terms would imply, the stereopsis said to be obligatory is a precise, valid, and subjectively certain experience of depth. Subjective depth intervals are directly related quantitatively to the disparity between the images in the two eyes. The results of the experiments just described indicate that, for obligatory stereopsis to occur, a simultaneous binocular perception of the objects involved must occur, in that horizontally associated disparate retinal elements must be stimulated or be active at the same time. This conclusion implies that true stereoscopic depth perception must rest on a physiological basis and is subject to a neuro-anatomic organization of the two retinas with the brain. The results also suggest that this organization may not demand a strictly *horizontal* association, for the "areas of stereoscopic vision" may also have within small limits a vertical extent. That is, a patent stereoscopic depth may be experienced from horizontally disparate images, even if these images are also slightly vertically disparate.

A facultative depth perception from disparate images may be said to be in the category of the Gestalt perceptions. Its occurrence is not limited by specifically horizontally associated disparate retinal elements. Its subjective counterpart seems confined mostly to a simple farther or nearer experience relative to the fixation point. It is certainly a vaguer experience of visual depth. This perception may arise from an unconscious comparison of distances between the two half-images of a given object and the images of the fixation point. When those distances are large, even this perception becomes less evident. The facultative depth perception is less evident with continuously seen images without eye movements, rather than when briefly seen. Eye movements then from one object to another will enhance this facultative experience in depth, especially in surroundings where the binocular parallax empirical clues (**2**, p. 139) are prominent. In classifying this mode of depth perception, one would be led to believe it is empirical and depends on the experience of the individual.

The results reported by van der Meulen and van Dooremaal would probably be explained on the basis of the facultative depth experience. Certainly on the basis of the results of the present experiment the inference that all of stereoscopic depth perception belongs in the same category must be rejected.

## Summary

Experiments have been described and data presented which clearly suggest that there are two aspects of stereoscopic depth perception: (*a*) an obligatory or quantitative sense of depth that necessitates the simultaneous stimulation of horizontally associated disparate retinal elements, and (*b*) a vaguer qualitative or facultative percept of depth also arising from disparate images but not dependent on specifically associated retinal elements. The obligatory stereoscopic depth is a true stereoscopic sensation of depth—precise, valid, that is to say, the depth interval is related to the magnitude of the disparity within the limits of a neuro-anatomic organization of the two retinas and the brain. The facultative depth experience, on the other hand, seems confined only to a "farther" or "nearer" percept relative to the point fixated. The response would be of the same nature as that to one of the empirical clues or motives for depth perception. On the basis of the present experiment, a true stereoscopic depth would be said to rest on a physiological basis limited by the neuro-anatomic organization of the retinas with the brain.

## References

1. Hofmann, F. B. Physiologische Optik (Raumsinn). In A. Graefe & T. Saemisch (Eds.), *Handbuch der gesamten Augenheilkunde*. Vol. 3. Part 2. Berlin: Julius Springer, 1925.
2. Ogle, K. N. *Researches in binocular vision*. Philadelphia: W. B. Saunders, 1950.
3. Ogle, K. N. On the limits of stereoscopic vision. *J. exp. Psychol.*, 1952, **44**, 253–259.
4. Tausch, R. Die beidaugige Raumwahrnehmung. *Z. exp. angewand. Psychol.*, 1953, **3**, 394–421.
5. van der Meulen, S. C., & van Dooremaal, T. C. Stereoscopisches Sehen ohne correspondirende Halbbilder. *Arch. Ophthal.*, 1873, **19**, 137–141.

(Received February 15, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# VISUAL AND AUDITORY CLOSURE[1]

BENJAMIN W. WHITE

*The George Washington University*

This study is an attempt to discover if certain abilities which have been discovered in visual perception can be found in auditory perception. The abilities are those Thurstone (12) has isolated in factorial studies and has labelled "Closure 1" and "Closure 2." So far these abilities have been demonstrated only in visual perception, though in Karlin's factorial study of audition (3) there is suggestive evidence that some sort of closure ability is present in this modality.

In gestalt theory, the principles of perceptual organization, among which closure is one, have a unique status. Koffka (6) makes it clear that they are basic to the understanding not only of perception, but also of other areas such as memory and learning. Within the province of perception, it is assumed that the laws of proximity, similarity, good continuation, etc. are valid in all sense modalities (6, p. 303). Thurstone's studies have shown that there are stable individual differences on tests requiring analysis and synthesis of visual material, tests such as his modification of the Gottschaldt figures and the Street Gestalt figures. If such differences reflect basic characteristics of perceptual organization, and if the gestalt assumption that the laws of perceptual organization are not specific to one sense modality is correct, it should be possible to construct tests in another sense modality which tap the same function and to demonstrate significant correlation between these and the visual measures.

In this study four of Thurstone's visual tests were employed: two tests with high loadings on Closure 1, and two with high loadings on Closure 2. The author then constructed auditory tests which on a priori considerations appeared to present *S* with the same tasks in interpreting auditory material as the visual closure tests. It was assumed that if an auditory test showed as high a correlation with a visual test as the latter did with the other visual test of the same factor, this could be taken as evidence that the factor was common to more than one modality.

## METHOD

*The test battery.*—The two tests of Closure 1 were Incomplete Words (14) and Mutilated Words Test, Form 1950 A (12). In both these tests *S* has to identify printed words which have letters or parts of letters missing from them. Two auditory tests were constructed so as to present *S* with a similar task of identifying fragmentary auditory material. The first consisted of tape-recorded spoken words from which various consonant sounds had been dropped. For instance, for the word "doorstep" *S*s hear "door ep." After hearing each incomplete word, *S*s were instructed to identify the word and to write it on an answer sheet. Mistakes in spelling were not counted as errors, score being the number of words correctly identified on the 40-word list.

The second auditory test also consisted of spoken words recorded on tape through a high-pass filter adjusted to pass frequencies above 2400. The words so recorded sounded sharp and "tinny," the high frequency sounds coming through very clearly, and the low frequency sounds being practically inaudible. As on the other test, *S*s were instructed to identify the words and write them on an answer sheet. Scoring on this 50-item test was the same as that described above.

[1] This paper is based in part upon a Ph.D. dissertation submitted in December 1952 to the Department of Psychology at the University of Michigan. The writer is indebted to the members of the committee and to Dr. Edward L. Walker, chairman of the Committee, who gave generous aid and invaluable criticism.

Concealed Figures, Form A (12) and Copying, (11) were the two visual measures of Closure 2. These tests demand that *S* identify or reproduce a geometric figure in a deceptive or distracting context. To present an analogous task in the auditory modality, two tests were employed. The first, Hidden Tunes, consisted of a series of items in which simple 3-, 4-, or 5-note melody patterns were presented, followed after a pause by a longer melody. The *S*'s task was to indicate by checking "yes" or "no" on his answer sheet whether or not the simple melody was contained in the more complicated melody. The musical instrument used in making the tape recording of this test was a Hammond organ. With this instrument it was possible to assure equal intensity for all notes in the melodic patterns. A switching device driven by a synchronous motor was attached to the speaker of the organ. This broke the circuit four times a second, thus making it possible to control the duration of each note played and the time intervals between notes. There were 50 such yes-no items in the test and the score was the number correct. Directions and practice items were recorded on the same tape.

The second auditory test was a tape transcription of Ability to Hear Words in Noise, Series L, Type W (4). This test was developed at the Psycho-Acoustic Laboratory, Harvard University, during World War II as an instrument to aid in the selection of communication personnel. The test is issued by the Laboratory on 12-in., 78-rpm records, and contains 200 items. Because of time limitations in this study, only the first 50 items were employed, and these were recorded on tape from the records. In this test *S* hears a series of common English words against a background of simulated airplane noise. His task is to write down on his answer sheet his best guess as to what each word is. The score is the total number of words correctly identified. Mistakes in spelling are not counted as errors.

IQ scores were also obtained for all *S*s whose school records contained this information. These IQ's were based upon the California Test of Mental Maturity.

*Subjects and Procedure.*—The *S*s were 116 students enrolled in two Michigan secondary schools, 72 boys and 44 girls. Mean age of the group was 15.4 yr. with an *SD* of 2.6 yr. The *S*s took the battery of tests in their classrooms during regular school hours. The *E* explained that the tests were being given for research purposes only and that since they were experimental measures, no interpretations could be given to individual scores, but that after the results had been analyzed, he would report general results back to *S*s. The testing required slightly more than one class hour. The *S*s found the tests, especially the auditory ones, interesting to take. They filled out a face sheet giving age, sex, and information concerning visual and auditory defects and extent of musical training.

The auditory tests were administered first. All were recorded on one spool of tape together with directions and practice items. Recording and playback were done on a Wilcox-Gay Recorder, Model 2A11, with a tape speed of 3.75 in. per sec. For playback in the classroom, an 8-in. speaker taken from a Bogan transcription player was attached to the tape recorder and placed in the front and center of the classroom.

## Results

*Distribution statistics.*—Table 1 summarizes the distribution statistics of the tests. Raw score distributions in most of the tests were reasonably symmetrical, though on one test, Mutilated Words, there was some skewing. One test, Distorted Words, was unduly difficult, the mean score being 6.9 out of 50 items. The reported reliability figures are based upon Kuder-Richardson Formula 20 (7). This formula is appropriate for the auditory tests since they were not speeded. For the Concealed Figures test the coefficient is based only upon the first 84 items of the test, which virtually all *S*s attempted. For Copying, a very heavily speeded test, a K-R reliability coefficient was inappropriate, so the figure reported is the test-retest reliability reported by MacQuarrie (8). The remaining visual tests, Incomplete Words and Mutilated Words, though speeded, have time limits which permit all *S*s an opportunity to "attempt" every item, so the use of the Kuder-Richardson formula is permissible.

The auditory tests have somewhat less satisfactory reliability than the visual tests, but were judged to have sufficient reliability for the purpose of this investigation.

*Test intercorrelations.*—The design of this study calls for an analysis of the correlations between these per-

TABLE 1

VISUAL AND AUDITORY TEST STATISTICS

| Tests | *N* | Mean | *SD* | Median | Range | Reliability | Correlation with IQ |
|---|---|---|---|---|---|---|---|
| *Visual* | | | | | | | |
| 1. Concealed Figures | 105 | 62.6 | 30.0 | 60 | −2–134 | .90[1] | .59 |
| 2. Copying | 68 | 11.7 | 5.4 | 12 | 1–29 | .86[2] | .51 |
| 3. Incomplete Words | 104 | 15.3 | 8.8 | 14 | 1–46 | .92 | .63 |
| 4. Mutilated Words | 104 | 19.5 | 7.3 | 10 | 5–40 | .87 | .56 |
| *Auditory* | | | | | | | |
| 5. Hidden Tunes | 111 | 37.8 | 6.7 | 38 | 24–50 | .83 | .39 |
| 6. Words in Noise | 110 | 19.9 | 5.2 | 20 | 6–35 | .67 | .26 |
| 7. Incomplete Words | 111 | 21.3 | 5.0 | 21 | 12–34 | .74 | .40 |
| 8. Distorted Words | 67 | 6.9 | 3.7 | 6 | 1–20 | .65 | .13 |

[1] Reliability based on first 84 items, which almost all *Ss* attempted.
[2] Test-retest reliability from MacQuarrie (8).

ceptual tests in order to determine if the abilities tapped by the visual tests can be tapped by the auditory tests as well. With this test battery, composed of four visual and four auditory tests, what sort of correlation pattern could be taken as support for the supposition that either the first or second closure factor exists in audition as well as in vision? The reader will recall that the battery of visual tests includes two measures of each closure factor. If an auditory test should correlate with both visual tests of a closure factor as highly as these two visual tests correlate with each other, this would be evidence that all three tests tap the same underlying trait. In such an event, it would be possible to say that this factor is not specific to one sense modality.

To take an example, this battery includes two visual tests with known high loadings on the first closure factor: Incomplete Words and Mutilated Words. Because both these tests have high loadings on the same factor, it is reasonable to expect that they will show a rather high correlation with each other. An auditory test, Spoken Incomplete Words, was constructed to be as similar as possible to the visual test Incomplete Words. If results show that this test in the auditory modality has correlations with the two visual tests mentioned above which are as high as the correlations of these visual tests with each other, there would be reason to believe that Spoken Incomplete Words taps the same ability in perception as the two visual tests.

This is the approach utilized in the analysis of the correlation matrix shown in Table 2. We will examine first the correlations between tests in the same modality and then proceed to examine the cross-modality correlations to see if any of the auditory tests appear to be tapping the same ability as the visual tests.

For the visual tests, shown in the upper left quadrant of the matrix, it is obvious that all four of them have substantial correlations with each other, despite the fact that they are not all primarily loaded with the same factor. For example, one of the highest correlations is between Concealed Figures and Mutilated Words, which are tests of the first and second closure factors respectively. From the previous studies of Botzum (1) and Pemberton (9), it had been anticipated that the correlations of the tests of the same factor should be higher than the correlations between the tests of different factors. The cross-factor correlations obtained between visual tests in this study are higher than

those found by these authors. Botzum used college students for his *Ss*, and Pemberton used volunteers at the University of Chicago, most of whom were graduate students. These groups were older than the *Ss* used in this study, and probably represent a higher level of general intellectual ability. Perhaps with maturation the two closure abilities become more differentiated so that the correlation between the factors is lowered. For *Ss* of high school age, however, these results suggest that the two closure abilities are not readily distinguishable.

The pattern of correlations between the auditory tests, shown in the lower right quadrant of Table 2, indicates a rather striking split. The Hidden Tunes test appears to be unrelated to the other three auditory tests, while they in turn seem to have something in common. All three of these tests involve the identification of spoken words, which suggests that this common content may account for the rather high correlations between them.

The cross-modality correlations, or those in the upper right or lower left quadrants of Table 2, are of primary interest. Here again Hidden Tunes stands alone. It shows high correlation with all of the visual tests, in marked contrast to its low correlations with the other auditory tests. Its correlations with the two visual tests of the second closure factor do not differ significantly from the correlation of these visual tests with each other. Its highest correlation, .66, is with Concealed Figures, the visual test on which Hidden Tunes was patterned most closely. Whatever perceptual ability these tests represent, it appears to be manifest in both modalities.

The other three auditory tests, though they correlate with each other, do not show any striking correlation with the visual tests. Tests of the significance of difference between correlations indicate that the correlations of these three auditory tests with the visual tests all differ significantly ($p < .01$) from $r = .65$ and $r = .74$, the correlations of the tests of Closure 1 and Closure 2 respectively. Thus, though these three verbal auditory tests show moderate correlation with each other, there is little evidence that this ability to identify spoken words under various conditions of omission, distortion, and distraction is related to either of the closure factors measured by the visual tests in this battery.

As the final column in Table 1

TABLE 2

INTERCORRELATIONS OF VISUAL AND AUDITORY TESTS

| Tests | Visual Tests | | | | Auditory Tests | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *Visual* | | | | | | | | |
| 1. Concealed Figures | | .65 | .57 | .69 | .66 | .10 | .20 | .27 |
| 2. Copying | .65 | | .44 | .50 | .54 | .03 | .28 | .21 |
| 3. Incomplete Words | .57 | .44 | | .74 | .55 | .22 | .37 | .14 |
| 4. Mutilated Words | .69 | .50 | .74 | | .57 | .21 | .29 | .20 |
| *Auditory* | | | | | | | | |
| 5. Hidden Tunes | .66 | .54 | .55 | .57 | | .00 | .23 | −.01 |
| 6. Words in Noise | .10 | .03 | .22 | .21 | .00 | | .51 | .49 |
| 7. Incomplete Words | .20 | .28 | .37 | .29 | .23 | .51 | | .40 |
| 8. Distorted Words | .27 | .21 | .14 | .20 | −.01 | .49 | .40 | |

Note.—All correlations are Pearson product-moment, and are computed using all *Ss* who took any pair of tests. Correlations were also computed using only *Ss* who had taken the entire battery. None of these differed from the coefficients reported above by more than .02.

indicates, several of these tests show rather high correlations with IQ. All the visual tests correlate with IQ above .50. In order to ascertain if the relationships described above would be altered if the variance attributable to IQ were taken out, correlations between the experimental tests were computed with IQ partialled out. These partial correlations are shown to the left of the diagonal in Table 3.

The essential features of the matrix remain unchanged. The visual tests still show substantial correlations with each other, and their cleavage into two pairs according to their known primary factor loadings is no more obvious here than it was in the original matrix in Table 2. The cleavage between the auditory tests is, if anything, more sharp than before, with the Hidden Tunes test unrelated to the other auditory tests. These three verbal auditory tests, however, maintain their high positive relation to each other. The cross-modality correlations are also essentially unchanged, with Hidden Tunes alone showing significant correlations with the visual tests. It is obvious from this analysis that the correlations between the Hidden Tunes test and the visual tests are not solely a function of intelligence, since the correlations are substantial even with this source of variance partialled out.

The chief concern of this study is with a hypothetical underlying ability or trait which can account for the obtained correlations between an auditory test and four visual tests. For this reason it is of interest to estimate what correlation could be expected between these tests if they were essentially pure and without error. As has been noted previously, the auditory tests are not very reliable by the conventional standards set for mental tests. To the right of the diagonal in Table 3 are shown the correlations between the visual and auditory tests corrected for attenuation. From these correlations it is possible to estimate the relationships which would be obtained if the tests were all perfectly reliable.

Since the visual tests have high reliability, this correction raises their intercorrelations only six to ten points. Among the auditory tests, however, the correlations between the three verbal tests are raised considerably, almost 20 points. The Hidden Tunes test is more isolated than before from the other auditory tests. The cross-modality correlations are raised some-

TABLE 3

INTERCORRELATIONS OF THE VISUAL AND AUDITORY TESTS WITH INTELLIGENCE HELD CONSTANT BY PARTIAL CORRELATION AND AFTER CORRECTION FOR ATTENUATION

| Tests | Visual Tests | | | | Auditory Tests | | | |
|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
| *Visual* | | | | | | | | |
| 1. Concealed Figures | | .75 | .63 | .78 | .76 | .13 | .25 | .35 |
| 2. Copying | .50 | | .50 | .57 | .65 | .04 | .36 | .28 |
| 3. Incomplete Words | .32 | .18 | | .83 | .63 | .28 | .45 | .18 |
| 4. Mutilated Words | .54 | .29 | .60 | | .67 | .27 | .36 | .26 |
| *Auditory* | | | | | | | | |
| 5. Hidden Tunes | .58 | .43 | .43 | .46 | | .00 | .29 | −.01 |
| 6. Words in Noise | −.07 | −.12 | .08 | .08 | −.11 | | .72 | .73 |
| 7. Incomplete Words | −.05 | .10 | .17 | .09 | .09 | .46 | | .57 |
| 8. Distorted Words | .24 | .17 | .08 | .16 | −.07 | .48 | .38 | |

Note.—Coefficients to the left of the diagonal are with IQ partialled out; those to the right of the diagonal are after correction for attenuation.

what, but the only ones above .50 are between Hidden Tunes and the visual tests. As before, the highest correlation for Hidden Tunes is with Concealed Figures. This correlation indicates that this hypothetical trait or ability may account for more than half the nonerror variance in these measures.

## Discussion

This study has yielded evidence that the ability represented by the second closure factor can be identified both in visual and in auditory perception. This is the factor which Thurstone describes as "the ability to keep a configuration in mind in spite of distractors" (13, p. 4). The visual and auditory tests which show the highest cross-modality correlation, Concealed Figures and Hidden Tunes, both require the subject to identify a figure, geometric or musical, when it is embedded in a more complex figure. On both tests most errors appear on items for which "yes" is the correct answer. This means that most of the raw score variance is attributable to the ability to recognize a figure when it is put in an altered context. This sort of task seems to embody the same dimension Koffka (6) had in mind when he spoke of the resistance of a psychophysical process to change. He ventures the opinion that if such a trait existed, it should be manifested in certain reasoning problems requiring reorganization of the perceptual field for solution.

Actually factorial studies by Botzum (1) and Thurstone (12) have indicated that tests of the second closure factor are closely associated with certain reasoning tasks. Pemberton's study (9) indicates that high and low scoring *Ss* on the Concealed Figures test differ significantly in their responses to a temperament schedule. These findings suggest that whatever the second closure factor represents, it is found in a wide variety of test situations.

With such evidence, it becomes somewhat misleading to refer to this factor in purely perceptual terms, and desirable to describe it in terms which will be applicable to a wider variety of responses, and which will suggest other ways in which this trait might be manifested. As a first approximation to such a description, we might say that *individuals differ in the degree to which their psychological processes are subject to alteration by the presence of other psychological processes which are adjacent either in space or time.* More precise specification of the term psychological process is beyond the scope of this paper since the only purpose here is to conceptualize the second closure factor so as to include its diverse manifestations in auditory and visual perception, in reasoning, and in temperament, and then to see if such a redefinition suggests further areas of behavior which might logically be subsumed under it. The following paragraphs suggest other phenomena which appear to be related to this definition, and indicate ways in which this relationship could be investigated empirically.

Simultaneous contrast might be described as one instance of interference between psychological events adjacent in space. It is conventionally measured by the extent to which the brightness of an object (psychological process) is altered by a change in the brightness of the surrounding field (alteration by the presence of other psychological processes which are adjacent in space). In this connection, Klein's work (5) on "perceptual attitudes" is of interest. His results indicate that there is a correlation between performance on the Concealed Figures test and scores on a test of simultaneous contrast. This suggests that there is some justification for saying that individual differences in contrast may be another manifestation of the trait described above. The various perceptual constancies may also be viewed as instances of the degree to which an object's perceived characteristics are altered by the conditions in the surrounding field. Individual differences in constancy, if the above definition is adequate, should show positive correlation with performance on a test like Concealed Figures.

Interference between events adjacent in time is shown in the Hidden Tunes test employed in this study. The melodic patterns in this test were distributed in time as compared with the spatially distributed patterns of the Concealed Figures test. Guilford's study (2) of the

Memory subtest of the Seashore Tests of Musical Talent gives another illustration of the alteration of a psychological process by the presence of other psychological processes which are adjacent in time. Many learning phenomena may also legitimately be characterized as psychological processes distributed in time. It would be interesting to discover whether an individual's vulnerability to retroactive inhibition in a rote learning situation is but another manifestation of this underlying dimension. If it were, there should be positive correlation between performance on a test such as Concealed Figures or Hidden Tunes and the degree of retroactive inhibition shown in a rote learning task.

In a sense this discussion has wandered rather far from the immediate findings of the study. However, one of the purposes of this investigation was to see what credence could be given the gestalt position that the laws of perceptual organization are basic not only to perception, but to other areas of behavior as well. This investigation indicates that individual differences in one characteristic of perceptual organization can be found in both the visual and auditory modalities. Other studies have been cited as evidence that this same characteristic is manifested in tests of reasoning ability and temperament. The discussion has been an effort to define this characteristic more broadly and to suggest other areas of behavior in which it might be manifested. The hypotheses presented are all testable with existing techniques, and should empirical investigation support them, this trait could emerge as a valuable concept in the understanding and prediction of human behavior.

## Summary

This study was concerned with discovering whether the closure factors isolated by Thurstone are specific to vision or whether they can generalize to the auditory modality. One auditory test was found which correlated with two visual tests of the second closure factor as highly as the two visual tests correlated with each other. This finding suggests that the second closure factor may be a central trait of considerable importance in a wide variety of behavior. Supporting evidence was found in other studies reporting significant relationships between this closure factor and tests of reasoning and temperament.

A redefinition of this trait of ability was made in order to free it from its specifically perceptual connotations. Under this broadened definition, several other psychological phenomena—simultaneous contrast, constancy, and retroactive inhibition—were suggested as additional situations in which this trait might be manifested. Correlation of individual differences in reaction to these situations with performance on a test such as Concealed Figures would throw further light on the generality of this trait, and offer valuable data both for personality research and perceptual theory.

## References

1. Botzum, W. W. A factorial study of the reasoning and closure factors. *Psychometrika*, 1951, **16**, 361–386.
2. Guilford, J. P., & Hilton, R. A. Configurational properties of melodies. *J. exp. Psychol.*, 1933, **16**, 32–54.
3. Karlin, J. E. A factorial study of auditory perception. *Psychometrika*, 1942, **7**, 251–279.
4. Karlin, J. E., & Abrams, M. H. Auditory tests of the ability to hear speech in noise. (OSRD, 1944; Publ. Bd., No. 22847). Washington, D. C.: U. S. Dep. Commerce, 1946.
5. Klein, G. S. The personal world through perception. In R. A. Blake & G. V. Ramsey (Eds.), *Perception: an approach to personality*. New York: Ronald, 1951. Pp. 328–355.
6. Koffka, K. *Principles of Gestalt psychology*. New York: Harcourt, Brace, 1935.
7. Kuder, G. F., & Richardson, M. W. The theory of the estimation of test reliability. *Psychometrika*, 1937, **2**, 151–160.
8. MacQuarrie, T. W. A mechanical ability test. *J. Personnel Res.*, 1927, **5**, 329–337.
9. Pemberton, C. A study of the speed and flexibility of closure factors. Unpublished doctor's dissertation, Univ. of Chicago, 1952.
10. Thouless, R. H. Individual differences in phenomenal regression. *Brit. J. Psychol.*, 1932, **22**, 216–241.
11. Thurstone, L. L. Primary mental abilities. *Psychometric Monogr.*, 1938, No. 1.
12. Thurstone, L. L. *A factorial study of perception*. Chicago: Univ. Chicago Press, 1944.
13. Thurstone, L. L. *Experimental studies of temperament*. Chicago: Psychometric Laboratory, Univ. of Chicago, 1951, No. 67.
14. Thurstone, L. L., & Thurstone, T. G. Factorial studies of intelligence. *Psychometric Monogr.*, 1941, No. 2.

(Received December 28, 1953)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# DISCRIMINATION TRANSFER ALONG A PITCH CONTINUUM[1]

ROBERT A. BAKER[2] AND STANLEY W. OSGOOD[3]

*Massachusetts Institute of Technology*

Improvement of pitch discrimination with practice has frequently been observed. While studies such as that by Wyatt (10) have demonstrated improvability, the factors responsible for the improvement and the most effective training procedures have not been fully determined (1, 3, 8). It is of considerable interest, in the latter instance, that Lawrence (7) has reported that learning was most efficient when his animals approached a difficult discrimination through a series of discriminations of graduated difficulty. This result is in conflict with the usual expectations for transfer of training, since it is normally assumed that for a fixed amount of practice on a training discrimination the transfer effect to a test discrimination will be greatest when the training and test discriminations are identical. The implications of the Lawrence finding are not only of great practical importance as regards training procedures but are of immediate theoretical importance for theories of discrimination learning. It, therefore, seemed advisable to repeat the Lawrence study utilizing human *Ss* and another discrimination continuum. Such a procedure, it was believed, would not only offer a better test of the general applicability of Lawrence's findings but would also furnish additional information regarding the most effective procedures for improving pitch discrimination.

## METHOD

*Experimental design.*—The experimental design was, essentially, the same as that employed by Lawrence (7). Four groups of *Ss*, ten to a group, were trained on pairs of tones differing only in frequency. All groups received the same number of reinforced trials of training on a given discrimination and were then tested on a difficult series. One group was trained throughout on the test series; a second was trained on a very easy discrimination and then shifted to the test; a third was trained first on a very easy discrimination but approached the test discrimination through a series of graduated steps; and the fourth group received no practice and served as a control. The amount of practice was held constant for the three experimental groups. All groups were given the difficult discrimination test prior to training. The differences between the pretest and the posttest scores thus furnished a measure of the effectiveness of the training. An outline of the experimental design, as well as the stimuli that were discriminated, is presented in Table 1.

*Subjects.*—The *Ss* were 40 male Army inductees. The average age was 22 yr., and all *Ss* were free of hearing defects as determined by Army medical standards. Scores on the Pitch Discrimination Test of the Seashore Measures of Musical Talent were used to assign *Ss* to one of the four groups. An attempt was made to cover the range of pitch-discrimination ability by selecting four *Ss* from each decile of the Seashore

[1] The research in this document was supported jointly by the Army, Navy, and Air Force under contract with the Massachusetts Institute of Technology. The authors wish to express their thanks and appreciation to Mrs. Lois K. Anderson and Mrs. Janet Zurier, who assisted in the collection and analysis of the data, and to Mr. Edward E. Caldwell for his assistance in assembling the electronic equipment. Thanks are also due to Dr. James W. Degan, who advised and assisted the authors on some of the statistical problems.

[2] Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology. Present address: Human Research Unit No. 1, The Armored Center, Fort Knox, Ky.

[3] Staff Member, Lincoln Laboratory, Massachusetts Institute of Technology, and Research Psychologist, U. S. Navy Electronics Laboratory, San Diego, California. Present address: Winchester, Massachusetts.

TABLE 1

EXPERIMENTAL DESIGN AND STIMULI VALUES
($N$ = 40 *Ss*, 10 *Ss* in each group)

| Group | Pretest | Practice Sessions | | | Posttest | Stimuli | cps Difference |
|---|---|---|---|---|---|---|---|
| | | 1 | 2 | 3 | | | |
| 1 | $D_1$ | $A_1$ | $A_2$ | $A_3$ | $D_1$ | A | 44–36 |
| 2 | $D_1$ | $A_1$ | B | C | $D_1$ | B | 34–26 |
| 3 | $D_1$ | $D_2$ | $D_3$ | $D_4$ | $D_1$ | C | 24–16 |
| 4 | $D_1$ | —— | —— | —— | $D_1$ | D | 14–6 |

Note.—The numerical subscript indicates the same test was employed with the exception that the order of presentation of the items was varied.

norms and assigning one of the four *Ss* to each group. From this procedure four groups covering the range of pitch discrimination and equivalent in initial ability, according to the Seashore scale, were obtained. The mean Seashore scores for the four groups were 38.5, 39.1, 37.7, and 38.7 items correct out of a possible 50 items. The *SD*'s for the four groups were 6.50, 6.70, 6.62, and 6.63, respectively.

Since the experiment was so designed that all *Ss* received the pretest on the difficult discrimination task, a second comparison of initial ability was available. The mean pretest scores for the four groups were 25.8, 21.9, 24.1, and 23.5 items correct out of a possible 40 items. The *SD*'s were 3.29, 3.80, 4.90, and 3.85, respectively.

*Apparatus.*—The stimuli consisted of pairs of pure tones equated in duration and intensity. The pairs of tones were always heard against a noise background. This noise served to mask any external sound sources normally present in the laboratory and, in addition, simulated many typical audio-communication systems wherein a signal is heard against a noise background.

The standard 1000-cps tone was produced by a Hewlett-Packard "Low Frequency Standard" Oscillator, Model 100D. The variable tone was produced by a General Radio Oscillator, Type 1304A. This instrument allows one to vary the signal in 1-cps steps. A laboratory type random noise generator was used for the background noise, and fed directly into three matched pairs of Cannon 2000-ohm headphones. The outputs of the two sine-wave oscillators were switched in parallel with the noise in a time sequence and were also led into the earphones. The "clickless" switching circuit passed one tone for .5 sec., allowed an interval of silence for 2.0 sec., and then passed the second tone for .5 sec.

The intensity of each of the pure tones, measured at the earphones, was +1 db re 1.0 volt. The signal was approximately 50 db above the noise measured in a 1-cycle band near the signal frequency. Thus the signal was comfortably audible for all listeners and the masking noise had little effect on frequency discrimination (4).

An oscilloscope on the Low Frequency Standard enabled *E* to monitor the output of both oscillators and to calibrate the frequency differences at the moment each pair of tones was presented.

*Procedure.*—After *Ss* had been assigned to the four groups, they were administered the 40-trial pretest on Stimuli D and were then trained according to the experimental schedule of Table 1. At the end of the practice periods all groups were then tested again on Stimuli D.

Of the 40 trials in each practice and test session half of the pairs of tones were identical 1000-cps tones. The other 20 pairs differed only in frequency. The frequency differences in each session covered a range of 8 cps in 2-cps steps. Thus, there were five steps in all. From Table 1 it can be seen that there was no overlap along the frequency continuum. In a given session each of the five steps occurred a total of four times; half of these at a given step occurred below the 1000-cps standard and half were above.

In each of the practice sessions the order of presentation of the same and different pairs was randomized with the restriction that no more than three pairs of a given type occurred successively. A switching circuit enabled *E* to present either the standard or the variable tone as the first member of a given pair. In each session, for half of the items, the standard tone appeared first and for the other half the variable tone preceded.

The *Ss* were tested and trained in an experimental room isolated from the equipment. The *E* communicated with *S* over a two-way intercom system.

At the beginning of the experiment *S* was

brought into the experimental room and seated at a small table facing two clearly labelled toggle switches and a red and a green panel light. The *S*'s switches controlled red and green panel lights in front of *E* enabling him to record *S*'s response. At the same time *E*, by throwing one of the two switches in the equipment room, was able to reinforce *S*. At the outset each *S* was instructed to listen carefully to each pair of tones and report *same* or *different* by throwing the appropriate switch. Three practice pairs, two of the pairs differing by 40 and −40 cps and an identical third pair, were administered before each practice and test session. The *S* was then given the 40-item pretest. Approximately 30 sec. elapsed between the presentations of each pair of tones. Following the pretest *S* was allowed to rest. At the end of 15 min. *S* was given the three warm-up pairs followed by the practice session. Fifteen minutes also elapsed between each of the practice sessions and between the last practice session and the posttest. During each of the practice sessions *S*'s responses were reinforced by *E*. No reinforcement, however, was given during the pretests or during the posttests.

In order to measure the effects of the training over a period of time and to insure that some learning resulted, each *S* repeated the entire pretest, training, and posttest procedure 24 hr. later.

## Results

Although the four groups of *S*s were successfully matched on the Seashore scale, which correlated .36 with pretest scores, the possibility existed that the groups differed significantly on the pretest. A non-significant *F* of 1.45, however, was obtained from the between-group variance analysis of the pretest scores.

The means and *SD*'s of the error scores for the four groups on all of the training and test sessions are presented in Table 2. It can be seen that only the gradual transition group (Group 2) revealed a significant improvement with training. The reduction in errors for this group was proportionately greater than that of the other groups at the end of the first day and strikingly greater at the end of the second day of practice. It should also be noted that the error scores for the experimental groups during training increased with the difficulty of the

TABLE 2

Average Errors for Training and Test Sessions
($N$ = 10 *S*s in each group)

| Test | Group 1 | | Group 2 | | Group 3 | | Group 4 | |
|---|---|---|---|---|---|---|---|---|
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| *Day 1* | | | | | | | | |
| Pretest | 14.2 | 3.29 | 18.1 | 3.80 | 15.9 | 4.90 | 16.5 | 3.85 |
| Practice | | | | | | | | |
| 1 | 4.0 | 2.65 | 3.8 | 2.22 | 12.3 | 4.24 | | |
| 2 | 3.2 | 2.74 | 4.1 | 4.18 | 11.7 | 3.95 | | |
| 3 | 3.4 | 2.68 | 6.0 | 5.00 | 14.5 | 3.96 | | |
| Posttest | 15.9 | 3.15 | 15.1 | 4.75 | 14.2 | 5.31 | 14.5 | 3.55 |
| *Day 2* | | | | | | | | |
| Pretest | 16.7 | 2.90 | 17.0 | 4.89 | 14.9 | 2.91 | 15.5 | 4.94 |
| Practice | | | | | | | | |
| 1 | 2.0 | 2.45 | 2.1 | 2.02 | 12.7 | 5.10 | | |
| 2 | 2.5 | 2.42 | 2.8 | 3.15 | 14.2 | 5.91 | | |
| 3 | 3.2 | 2.71 | 6.1 | 5.38 | 12.9 | 5.26 | | |
| Posttest | 16.5 | 2.33 | 12.0 | 4.64 | 14.4 | 4.54 | 15.7 | 4.51 |

TABLE 3

ANALYSIS OF COVARIANCE FOR ERROR SCORES ON POSTTEST

| Source | Σ (Errors of Estimate)² | df | Mean Square | F |
|---|---|---|---|---|
| Total | 855.83 | 38 | | |
| Within groups | 675.13 | 35 | 19.29 | |
| Between adjusted means | 180.70 | 3 | 60.23 | 3.12* |

* $.05 > p > .01$.

discriminanda. Thus it is evident that the stimuli employed formed a series of progressively greater difficulty and that a "true" discrimination continuum existed.

Although the groups were matched in terms of the Seashore scale, their mean error scores on the initial pretest were slightly diverse. Therefore, the error scores of the final posttest were subjected to an $F$ test involving covariance. The results of this analysis are presented in Table 3. It can be seen that an $F$ of 3.12, significant between the .01 and .05 levels, was obtained. Thus, it is indicated that the differences in the group means on the final posttest cannot be accounted for by differences in the mean level of initial ability as measured by the initial pretest.

Since the groups were initially matched $S$ for $S$ on the Seashore scale, an evaluation of the improvement within and between the various groups is of some interest. The $t$'s between the initial pretest and final posttest scores for each group revealed that the gains of the gradual transition group were significant at the .01 level of confidence. None of the other differences was statistically significant.

Owing to the possibility that the improvement of the gradual transition group resulted from improved recognition of the identical tone pairs rather than a true change in the discrimination threshold, an analysis of their errors was made. On the initial pretest trial 49% of the total errors of the group occurred for the pairs of tones differing in frequency. On the final posttest trial, however, only 47% of the errors occurred for the pairs differing in frequency. Therefore, some improvement in the direction of a lower threshold of discrimination is indicated.

With regard to the intergroup comparisons, a $t$ of 2.23, significant at the .05 level, was obtained from a comparison of the gradually trained and control groups. An even larger ratio ($t = 3.81$) was obtained when the gradually trained group was compared with the group trained on the easy discriminanda. While the comparison of the gradually trained group with the group trained on the test items was not significant at the .05 level, the obtained $t$ of 1.90 was significant at the .073 level. None of the other intergroup comparisons approached significance.

A detailed comparison of group performances is presented in Fig. 1 where the scores made by each $S$ on the initial pretest and the final posttest

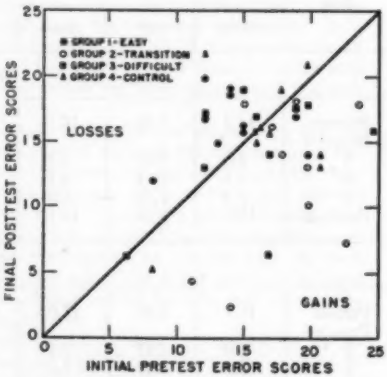


FIG. 1. Individual changes in error scores after training ($N = 40$)

are shown. Of particular interest are the trends shown by the groups. Thus it can be seen that for the easy discrimination group eight *S*s showed a loss in performance between pretest and posttest, and two showed gains. In the gradual transition group, however, only one *S* showed a loss and nine showed gains. For the other two groups it is clear that there were no evident trends.

The question might be raised as to whether the variable treated in the experiment is normally distributed. On the assumption that not too much is known about this, it was decided to use a distribution-free test to evaluate the significance of the changes in performance shown by Groups 1 and 2 and compare the results with those shown by the *t* tests. A nonparametric test especially suited to the purpose is the Sign Test (2). This test takes into account the direction of the change in performance from pretest to posttest, i.e., whether plus or minus, and makes no assumption as to normality.

The Sign Test evaluations confirmed the interpretations of group performance changes yielded by the *t* tests. For example, the performance shift for the gradual transition group was shown to be significant at the .02 level. The change in performance for the easy discrimination group was not significant ($p = .11$). As with the *t* tests, two-tailed tests were used throughout.

## Discussion

The results suggest that the most efficient way to establish a pitch discrimination is to approach the task in steps of gradually increasing difficulty. This conclusion is also in essential agreement with the findings of Lawrence (7), who employed animals and visual discriminanda. There was, however, one point of disagreement. From an inspection of Table 2 it is clear that the performance of the easy discrimination group (Group 1) on the test discrimination gradually deteriorated with practice. In this instance, it would seem that repeated experience with large stimulus differences increases the difficulty of detecting smaller differences. This result is more in keeping with other studies of transfer in which it has been found that transfer from the most dissimilar series is usually negative. An alternative explanation, however, should be considered. The process at work may be similar to what Helson (5) has summarized in the concept of *adaptation level*. In the present experiment the judgments of difference were made against the previous level of expectancy and thus were erroneous.

In spite of the fact that the performance decrement of Group 1 is in contradiction to Lawrence's results, the somewhat arbitrary nature of the stimulus differences involved, the employment of different stimulus continua and different species of *S*s, as well as other factors specific to this experiment, must be noted. A further possibility exists that training on a pitch discrimination of somewhat greater difficulty, for example Stimuli B or C of Table 1 instead of Stimuli A, would reveal a greater amount of transfer when tested on the difficult task.

Although the amount of practice and the number of reinforcements (one per trial) were identical, the number of successes varied between groups. During practice the easy discrimination group made the fewest errors. They were followed by the gradually trained group and the difficult discrimination group. While it is possible that the superiority of the gradually trained group was partly due to the fact that they received more positive reinforcements than the difficult group during practice, with some resulting facilitation of learning, the fact that the easy discrimination group received even more positive reinforcement suggests that reinforcement is not the only factor involved.

These findings, it is believed, are of more than passing interest to those concerned with pitch-discrimination training. For the present task the superiority of the gradual transition procedure is quite evident. The efficacy of such a training procedure is not only attested by the present results and those of Lawrence (7), but is also supported by several observations in the literature (6, 9). Despite the fact that generalizations from the present study to all pitch-discrimination tasks are severely limited, it is possible that extended training in the present task would not only produce greater improvement but would also improve performance on similar discriminanda such as sonar signals. Further investigation of the present findings is planned.

## Summary

Four groups of ten *Ss* were trained on pairs of tones differing only in frequency. All groups received an equal number of reinforced training trials on a given discrimination and were subsequently tested on a difficult series. One group was trained throughout on the test series; a second was trained on an easy discrimination and then shifted to the test; a third approached the test discrimination through a series of graduated steps; and a fourth group received no practice and served as a control. All groups were administered the difficult discrimination test prior to training. When these groups were compared in terms of improvement on the difficult discrimination test, it was found:

1. The only significant evidence of learning was obtained for *Ss* who approached the test discrimination through a series of graduated steps.

2. Performance on the test discrimination appeared to deteriorate for those *Ss* trained on the very easy discrimination and abruptly shifted to the test discrimination. The decrement in performance was not, however, statistically significant.

## References

1. Chapanis, A., Garner, W. R., & Morgan, C. T. *Applied experimental psychology.* New York: Wiley, 1949.
2. Dixon, W. J., & Mood, A. M. The statistical sign test. *J. Amer. statist. Ass.*, 1946, **41**, 557–566.
3. Garner, W. R. Auditory signals. In *NRC, Human factors in undersea warfare.* Washington, D. C.: National Research Council, 1949. Pp. 201–217.
4. Harris, J. D. Studies on pitch discrimination in masking. II. The effect of signal/noise differential. *J. acoust. Soc. Amer.*, 1947, **19**, 816–819.
5. Helson, H. Adaptation-level as a frame of reference for prediction of psychophysical data. *Amer. J. Psychol.*, 1947, **60**, 1–29.
6. Lashley, K. S. The mechanism of vision: XV. Preliminary studies of the rat's capacity for detail vision. *J. gen. Psychol.*, 1938, **18**, 123–193.
7. Lawrence, D. H. The transfer of a discrimination along a continuum. *J. comp. physiol. Psychol.*, 1952, **45**, 511–516.
8. Neff, W. D., & Thurlow, W. R. Auditory discrimination in sonar operation. In *NRC, Human factors in undersea warfare.* Washington, D. C.: National Research Council, 1949. Pp. 219–230.
9. Pavlov, I. P. *Conditioned reflexes.* Oxford: Oxford Univ. Press, 1927.
10. Wyatt, R. F. Improvability of pitch discrimination. *Psychol. Monogr.*, 1945, **58**, No. 2 (Whole No. 267).

(Received February 12, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# CONSPICUITY OF FLASHING LIGHT SIGNALS OF DIFFERENT FREQUENCY AND DURATION[1]

SIEGFRIED J. GERATHEWOHL

*USAF School of Aviation Medicine*

It has been shown in previous studies of signal conspicuity that flashing light signals are more conspicuous than steady ones under conditions of low brightness contrast (**3, 4, 5**). Conspicuity has to do with the efficacy of the light to attract attention. We have measured conspicuity by the speed with which *S* responds to a light stimulus. By a steady light we mean a signal which is uninterrupted and has a constant luminance; whereas a flashing light consists of alternate periods of light and dark that can be perceived by the eye. We are not concerned here with the comparative effectiveness of a single light flash. The experimental evidence on this question has been summarized by Long (**8**) and need not be repeated in this paper.

In the past, two measures have been used to express the comparative effectiveness of flashing light signals. These are the comparative intensities required either for threshold or for equal apparent brightness above threshold. The two measures agree in showing that the effectiveness of light stimuli of short duration is based on the reciprocal relation of intensity and time. In the experiments reported here, the conspicuity of flashing light signals outside of the Bunsen-Roscoe range is investigated by applying three different flash frequencies and flash durations.

[1] The author wishes to thank Mr. David Rubinstein, School of Aviation Medicine, Department of Biometrics, for help in the statistical evaluation of the data.

## METHOD

The test apparatus consisted of (*a*) flashing signals of various flash frequencies and durations; (*b*) a complex test situation to distract *S* from the test signals; and (*c*) timer and counters to record the speed with which *S* responded to the test signals. The present description will be brief, since a more complete one is given elsewhere (**3, 4, 5**).

The *S* was seated before a visual display consisting of two vertically arranged screens. The test signals appeared as spots of light in the center of the upper screen, at an angular height of about 15°30′ above the center of the lower screen, i.e., the central line of sight. Light signals used to distract *S* appeared as colored or patterned dots on the lower of the two screens, approximately at *S*'s eye level. In order to let *S* freely scan the whole stimulus field, his head was not fixed.

The test signals were produced by illuminating the upper screen from behind with two 110-v., 25-w. GE Mazda lamps. The luminance of the test signals was controlled by Wratten neutral gelatin filters. The intermittency of the signals was obtained by a cam-driven relay system which turned the test light on at given intervals. The incandescence and nigrescence times for the lamps were about 1/10 sec. Experiments were made successively at a high (74.2), a medium (1.0), and a low (.19) brightness contrast. Contrast is defined as $L_2$–$L_1/L_1$, where $L_2$ is the total luminance of the signal and $L_1$ is the background luminance.

At each contrast, flashing signals of 1, 2, and 4 flashes per second were tested, with durations of $\frac{1}{2}$, $\frac{1}{4}$, and $\frac{1}{8}$ sec., respectively. Thus the total luminous flux emitted per second was assumed to be equal in the three cases; thereby slight differences due to incandescence and nigrescence were not considered. The lower-frequency signal was always on the left side, the higher-frequency signal on the right side of the screen. Since the test signals never appeared together, they were treated as independent stimuli.

The luminance of the upper screen (background), on which the test signals appeared, varied from 2.0 to 2.8 mL. in various experi-

mental sessions. The luminance of the test signals was always matched while the lights were set steady. The signals were tested on nine groups of 25 or 26 *S*s each, with each of the possible pairings of flash frequencies (1–2, 1–4, 2–4) combined with each of the brightness contrasts (74.20, 1.00, .19).

The *S* was instructed to press the right or left key, mounted on the two response levers of a control box, whenever he was aware that a test signal had appeared on either the right or left side of the top screen. The time between the onset of the signal and *S*'s response was recorded by one of two Standard Electric time clocks. Special care was taken that the timer would not start in the "dark period" of intermittence, but only when the signal was present on the screen.

The apparatus used to present the distracting stimuli is of relatively little interest here. It consisted of the lower screen on which light signals were projected from behind. The luminance of the distracting stimuli varied from 11.9 to 119.6 mL. The *S* had to respond to these signals by moving one of the two handles of the control box. Two tone signals of different pitch were also used as distractors. To these *S* was instructed to respond by operating one or the other of two pedals.

The basic experimental series consisted of 270 distracting light and tone stimuli and 30 test signals, presented during a period of about 9 min. Each test series contained 15 test signals of a lower and 15 of a higher frequency. The series was repeated on the same *S* after a 3-min. rest period. The test signals were released automatically in an unpredictable sequence. Each signal was terminated when *S* pressed the proper key. If he failed to do so, the test signal would stay on until the same signal appeared again. Although there were only a few cases of overlapping of signals, the numbers of responses were not equal in all cases. Hence, the response time had to be computed from the individual counters for each type of signal at each of the three contrasts. Extreme cases of failure to respond were excluded from the experimental groups. Furthermore, some confusion occurred when *S* pressed the wrong key in response to the signal properly perceived. While this determined the design of the experiment, such cases would not radically bias its result.

Each clock recorded the total response time for the 15 presentations of either type of flashing light contained in a normal series. The average time for one response was then computed and expressed in milliseconds. This time, rather long because of the complexity of the test and the low brightness contrast, was used as an index of conspicuity for that particular light signal.

TABLE 1

MEANS AND *SD*'S FOR EACH TEST CONDITION*

| Flash | | *N* | Series 1 | | Series 2 | | Totals Mean |
|---|---|---|---|---|---|---|---|
| Frequency | Duration | | Mean | *SD* | Mean | *SD* | |
| Contrast .19 | | | | | | | |
| 1 | $\frac{1}{2}$ | 50 | 1122.3 | 42.1 | 926.4 | 26.3 | 1024.4 |
| 2 | $\frac{1}{4}$ | 50 | 1071.8 | 37.1 | 868.1 | 20.0 | 970.0 |
| 4 | $\frac{1}{8}$ | 50 | 1061.4 | 47.2 | 831.2 | 16.7 | 946.3 |
| Contrast 1.00 | | | | | | | |
| 1 | $\frac{1}{2}$ | 50 | 963.2 | 28.8 | 788.1 | 24.9 | 875.7 |
| 2 | $\frac{1}{4}$ | 50 | 950.1 | 25.6 | 792.3 | 15.8 | 871.2 |
| 4 | $\frac{1}{8}$ | 50 | 934.7 | 30.7 | 809.8 | 31.2 | 872.3 |
| Contrast 74.20 | | | | | | | |
| 1 | $\frac{1}{2}$ | 52 | 894.2 | 22.3 | 756.8 | 24.5 | 825.5 |
| 2 | $\frac{1}{4}$ | 50 | 872.7 | 21.0 | 731.0 | 16.5 | 801.9 |
| 4 | $\frac{1}{8}$ | 50 | 905.0 | 32.3 | 738.9 | 20.0 | 830.0 |

* The values given in this table are the combined means, in milliseconds, of each frequency seen by two groups of *S*s at each particular contrast. *N* is the number of *S*s viewing a particular signal.

TABLE 2

$p$ VALUES FROM 3-WAY ANALYSIS OF VARIANCE OF SUBJECTS (A), FREQUENCY (B), AND SERIES (C) IN NINE EXPERIMENTAL GROUPS

| Source | $p$ Values in Group | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 |
| A | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 | .046 | <.005 |
| B | .132 | .084 | .346 | .202 | .048 | >.50 | >.50 | >.50 | .294 |
| C | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 | <.005 |
| AB | >.50 | .195 | .214 | .180 | .326 | >.50 | .046 | >.50 | .455 |
| AC | .229 | .397 | .016 | .229 | .037 | .163 | .006 | >.50 | >.50 |
| BC | >.50 | .294 | .436 | .258 | .122 | .042 | >.50 | .246 | >.50 |

## RESULTS

A survey of the results is given in Table 1. The values in this table are the pooled means and *SD*'s of each frequency seen by two groups of *S*s at each brightness contrast. The entries for *N* refer to the number of *S*s who viewed the particular signal, indicating that each *S* was tested on two different signals. The results show that there is a consistent decrease of reaction time at the lowest brightness contrast with increasing flash frequency. However, at the medium and high contrasts response time seems to be relatively unaffected by the variation of frequency.

An analysis of variance was made for each of the nine groups. All of these nine analyses showed very significant differences between *S*s and between test series. The interactions were significant in some of the groups. Usually they were small when compared with the main effects of *S* or series. A survey of the $p$ values obtained from the nine groups is given in Table 2. The learning effect, which is expressed by the difference between the two test series, is statistically significant beyond the .005 level of confidence.

Finally, an analysis of variance was made on the effects of contrast and frequency. In this analysis the unpooled test series were used. They are the group means of the frequencies to which the group was exposed. Since each group saw two different signals on a particular contrast, 36 entries were thus obtained which were treated as separate blocks. The results of this analysis are given in Table 3. From Table 3 it appears that the increase in contrast definitely improves the conspicuity of the flashing light signal.

Performances for the different flash frequencies do not show such a consistent trend. It seems plausible to assume that at the higher brightness contrasts all flash frequencies are seen with about the same facility. However, at the .19 contrast the response time decreases as the flash frequency increases. Statistically significant differences for the flash frequency at

TABLE 3

ANALYSIS OF VARIANCE OF THE MAIN EFFECTS OF CONTRAST AND FREQUENCY IN ALL GROUPS COMBINED

| Source | *df* | Mean Square | *F* |
|---|---|---|---|
| Contrast | 2 | 48915414 | 6.987* |
| Frequency | 2 | 2793324 | .399 |
| Contrast × Frequency | 4 | 1124783 | .161 |
| Error | 27 | 7000808 | |
| Total | 35 | | |

* Significant at .01 level.

the .19 brightness contrast were established by a $t$ test which yielded a $p$ value of .02.

## Discussion

In 1885 Bloch demonstrated that, within a certain limited range of durations, the product of intensity and duration must be constant in order to maintain a constant visual effect (1, 7). Recently, Long (8) has shown that the "critical duration" indicating the point where the break of this relationship occurs, falls near .10 sec.; and this result closely approximates those obtained by other investigators (2, 8). When the exposure time of the stimulus is longer than this action time, the intensity is a constant. In terms of our problem we should expect that when light stimuli longer than .10 sec. are used, no differences in conspicuity will occur, provided that the luminance of the signals is the same. The surprising result of our study is the fact that, with low contrasts at least, signals made up of short flashes are more conspicuous than signals made up of long flashes. This finding seems to contradict the relation $I = C$ which was found to be valid by Long (8) when the exposure time was longer than .10 sec.

An attempt to explain this phenomenon has been made elsewhere (5). Suppose we postulate that, for large and medium contrasts, the immediate effect of the signal is so great that response occurs almost at once. In that case, it would be of small importance whether the light stays on for a short or a long period of time, or whether it blinks at a fast or at a slow rate. The finding that conspicuity appears to be somewhat greater for steady signals, and for signals of two flashes per second, at higher contrasts suggests that the immediate effect alone does not fully determine the response time. The data imply that there may be an optimal intensity-time relation which still remains to be found.

With low contrasts, short and fast flashing signals are more conspicuous than long and slow flashing ones. This also agrees with our former findings which indicate that at low contrasts flashing signals are more easily seen than steady ones. In other words, alternating the signal more frequently will increase its conspicuity. Hartline, Wagner, and Nichol (6) and Noell (9) have shown that there is a maximum of stimulation at the beginning of each signal. The more often such changes occur, the greater is their statistical effect of being seen. Davy (2) has shown that there is a certain degree of summation of energy so that the effect of the two flashes is greater than either alone, if the second flash falls within or shortly after the end of the critical duration. This may be the case with our shortest signal where the second flash follows .125 after the foregoing one. It is also suggestive through Davy's discussion of his results that the summation effect may be pronounced at the low contrast. As to the two other signals used in our study, the time interval is so long that no integration effects between sequential flashes can be expected.

Another possible interpretation of our results should be mentioned here. We may not be actually dealing with perceptual differences between the various types of signals. One can argue that the faster flashing signal is not really seen better, but that it may be a stronger agitator for the motor response after it has been detected.

It is to say, however, that an enhancement of apparent brightness occurs with flicker (Bartley effect), and that this is a purely perceptual phenomenon. This seems to be true in a similar way for flashing signals below the Bartley minimum. Hence, we may conclude that the differences in response time are caused by the ease or difficulty with which the signal is seen, rather than by a subsequent effect on the motor response. The main reason for our findings seems to be that we are dealing here with a series of flashes instead of with the effectiveness of a single flash.

## Summary

The experiments reported here have produced evidence that when *S* is engaged in a very complex psychomotor task, and does not know when and where a light signal may appear, its efficacy as a warning or indicator is determined not by the luminance of a single flash alone, but by the conspicuity of a series of flashes. The results suggest that, if the brightness contrast is 1.00 or 74.20, i.e., according to our previous findings, close to or larger than 1.00, *S* will respond to a series of light flashes in a complex situation with about the same speed whether the flash becomes visible only once each second for ½ sec., two times for ¼ sec., or whether it occurs four times per second for only ⅛ sec. At the low contrast, however, the fast flashing light of short duration seems to be more conspicuous than the slow flashing signal of a longer duration.

## References

1. Blondel, A., & Rey, J. The perception of lights of short duration at their range limits. *Trans. Illum. Engng Soc. N. Y.*, 1912, **7**, 625.
2. Davy, E. The intensity-time relation for multiple flashes of light in the peripheral retina. *J. opt. Soc. Amer.*, 1952, **42**, 937–941.
3. Gerathewohl, S. J. Conspicuity of flashing and steady light signals. I. Variation of contrast. *USAF Sch. Aviat. Med. Spec. Rep.*, April, 1951.
4. Gerathewohl, S. J. Conspicuity of flashing and steady light signals. II. High contrasts. *USAF Sch. Aviat. Med. Proj. Rep.*, 1952, No. 21-24-014, Rep. No. 2.
5. Gerathewohl, S. J. Conspicuity of steady and flashing light signals: variation of contrast. *J. opt. Soc. Amer.*, 1953, **43**, 567–571.
6. Hartline, H. K., Wagner, H. G., & Nichol, E. F., Jr. The peripheral origin of nervous activity in the visual system. *Cold Spring Harbor Symposia on Quantitative Biology*, **17**, 1952.
7. Lash, J. D., & Prideaux, G. F. Visibility of signal lights. *Illum. Engng*, 1943, **38**, 481–492.
8. Long, G. E. The effect of duration of onset and cessation of light flash on the intensity-time relation in the peripheral retina. *J. opt. Soc. Amer.*, 1951, **41**, 743–747.
9. Noell, W. K. Studies of the electrophysiology and the metabolism of the retina. *USAF Sch. Aviat. Med. Spec. Rep.*, October, 1953.

(Received January 27, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# THE ACQUISITION OF COMPOUND CONCEPTS AS A FUNCTION OF PREVIOUS TRAINING[1,2]

HOWARD H. KENDLER AND ROBERT VINEBERG[3]

*New York University*

The present study is one of a series concerned with an attempt to relate S–R theory to human reasoning behavior. The empirical problem investigated was a specific one related to concept formation behavior in a conventional card-sorting situation. It sought to determine the influence the learning of simple concepts had upon the learning of compound concepts which involved combinations of simple concepts. An example of a simple concept would require sorting a series of cards by the color of the figures on them. Another simple concept would require *S*s to sort the cards according to the size of figures. A compound concept would involve sorting on the basis of both size and color.

Our experimental design involved three groups, each required to learn two simple concepts successively. Following this, each group was required to learn the test concept, which was a compound concept. One group learned both of the simple concepts of which the test concept was composed, the second group learned only one, while the third group learned neither of these concepts. The expectation was that the rate of learning the test concept would be directly related to the number of simple concepts learned which were involved in the test concept. Since the learning of the test concept depended upon the development of two specific response tendencies, it would be expected that previous training involving these response tendencies would facilitate the acquisition of the test concept.

## METHOD

*Subjects.*—Male students from the general psychology course at University College of New York University were used. They were naive as to the nature of the experiment. The final results included the data from 62 *S*s. Four *S*s had to be discarded because of their failure to learn one of the simple concepts, while three *S*s were unable to complete their 192 test trials within the time limit of 1 hr. One each of these latter three *S*s came from each of the experimental groups of the second study.

*Procedure.*—The conventional card-sorting test procedure was used. The *S*s were required to sort the cards with respect to a group of stimulus cards. Correct responses were followed by *E*'s saying "right," and incorrect responses were called "wrong." The initial form of the N.Y.U. Card Sorting Test was used. This test was patterned after the Wisconsin Card Sorting Test (1). Our test consisted of 32 response cards which varied in terms of four forms (circle, crescent, square, and triangle), four colors (black, gray, yellow, and orange), and two sizes. The designs appeared on a 3-in. square card with the designs themselves being either 1 or 2 in. in height.

[1] Editor's note: This article was submitted originally on April 20, 1953, and in substantially the present form. It was returned to the authors for certain revisions; through an unfortunate accident the Editor delayed action on the revised manuscript for about nine months, and the present manuscript was not received until June 28, 1954. Therefore, this article is published out of turn, in terms of the stated date of receipt, in order to rectify in part the errors of the past.

[2] This experiment was performed in connection with Project NR 150-064, under Contract NONR-187(00) between the Office of Naval Research and New York University. The authors are indebted to Drs. Benton J. Underwood and Tracy S. Kendler for their helpful criticisms and suggestions.

[3] Now with the Training Research Division of the Human Resources Research Office, George Washington University.

Two studies, with the same experimental design, were conducted. In the first study, three groups were required to learn two simple concepts successively. The *S*s were seated individually in front of a table on which was a tray with two rows of four compartments in which the cards could be placed. In the two center compartments in the row farthest from *S* were placed two stimulus cards: a large yellow crescent and a small black square. The response cards could be sorted into two categories represented by these two stimulus cards in four different ways. The basis of such sorting would involve four simple concepts: size, color, shape, and "part-whole." Acquisition of the size concept would be indicated by *S*'s placing the cards having large designs under the large yellow crescent and small designs under the small black square. The color concept required *S* to sort chromatic cards (yellow and orange) under the yellow stimulus card, and achromatic cards (black and gray) under the black stimulus card. The shape concept involved placing curvilinear figures (crescent and circle) under the stimulus card with the crescent and placing rectilinear figures (square and triangle) under the stimulus card containing the square. The part-whole concept was based upon the idea that a triangle is part of a square and a crescent is part of a circle. Consequently, when *S* was learning this concept, he was required to sort part figures below the crescent and whole figures below the square.

At the beginning of the experiment, *S* was given the following instructions: "I want you to put each of these cards (pointing to the response deck) in the slot below the card (pointing to the stimulus cards) with which you think it belongs. After you place each card, I will tell you whether you are right or wrong. Your object is to get as many as possible right. Do you understand? Go ahead and try the first one."

After *S* reached a criterion of ten successive correct sorting responses, he received the following instructions: "That was very good. I would now like you to try a different problem. This one has a different solution." Each *S* was permitted a total of 5 decks (160 cards) to learn each of the simple concepts.

In the first study (represented by the letter A), Group 1A learned the shape concept and then the size concept; Group 2A learned the shape concept and then the color concept; and Group 3A learned initially the color concept and then the part-whole concept.

Following the learning of the two simple concepts, *S* was told: "That was fine. Now here is a third problem for you to try. Now we will have four categories."

Two additional stimulus cards were added; a large gray triangle appeared in the first compartment of the tray and a small orange circle was placed in the fourth compartment. The *S*s in all three groups were required to learn the same test concept, a compound concept involving both the shape and size concept (see Table 2). Examples of correct responses would be as follows: a response card with a small black crescent should be placed under the small orange circle because they are both small and have the same shape (the figures are composed of curved lines). A large orange square should be placed under the large gray triangle, because both are large and have straight lines. The criterion of learning for the test concept was also ten successive correct responses. Each *S* had a limit of six decks (192 cards) in which to achieve the criterion of learning. If *S* failed to learn the test concept within this 192 trial limit, he was considered as being unable to solve the test problem.

The second study (represented by the letter B) had a similar design and procedure (see Table 2). Three groups of *S*s were required to learn two simple concepts successively. Group 1B learned the shape and then the size concept, Group 2B learned the shape and color concepts, while Group 3B learned the color and part-whole concepts. The major difference between the first and second study was the nature of the test concept. In the second study the test concept, which was also a compound concept, involved both the color and part-whole concepts. According to such a concept, a correct sorting of a response card with a small black crescent would involve placing that card under the stimulus card having a large gray triangle, since both these cards have achromatic coloring, and both are part figures.

The inclusion of the second study in our experimental design was to control for the possibility that the rate of acquisition in the first study would be a function of some general effect induced by the learning of the simple concepts rather than due to the specific interaction *between* the simple concepts and the test concept. For example, if Groups 1A and 1B learned the test concept most rapidly, then we would attribute such a finding to some general facilitating effect the learning of the simple concepts of size and shape had upon the learning of all compound concepts. If, on the other hand, Groups 1A and 3B learned the test concept most rapidly, we would ascribe such results to the interaction between the successive learning of the simple concepts and the test concept.

## Results

Table 1 reports the data associated with the learning of the training concepts in both studies. None of the differences between the groups with the same training experience (e.g., Group 1A and 1B) was significant at the .05 level.

The results of the first study are presented in Table 2. Two *Ss* in Group 3A failed to learn the test concept within the allotted 192 test trials. These *Ss* were assigned scores of 192, and the calculation of both the mean and *SD* of Group 3A included these two unsuccessful *Ss*. It was felt that such a procedure was defensible since it did not exaggerate the differences between the groups.

It was deemed advisable, because of the problem raised by the two unsuccessful *Ss* in Group 3A as well as the large differences between the variance of Group 3A and the other two groups, to use a simple nonparametric test to evaluate the significance of the differences between the various experimental groups. *U* tests (4) revealed that the differences between Groups 1A and 2A, 2A and 3A, and 1A and 3A were significant, on the basis of a one-tailed hypothesis, at the .02, .05, and .01 levels, respectively. The probability that these three comparisons would result in three significant differences beyond the .05 level is .0025.

Table 3 reports the data of the second study. Only 8 of the 31 *Ss* in this study were capable of solving the test concept within the allotted number of trials. The results, however, do suggest a trend; the percentage of *Ss* solving the test concept, as well as the number of trials required by these successful *Ss* to reach the criterion of learning, appears to be related to the number of simple concepts appropriate to the test concept which had previously been learned.

## Discussion

The results of both studies appear to be, in a general way, consistent with the prediction made that the rate of learning the test concept would be directly related to the number of simple concepts appropriate to the test concept which had been learned. In the first study a very definite and significant relationship was

TABLE 1

Mean Number of Trials Required to Attain Training Concepts

| Group | Shape | | Size | | Group | Shape | | Color | | Group | Color | | Part-Whole | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | M | *SD* | M | *SD* | | M | *SD* | M | *SD* | | M | *SD* | M | *SD* |
| 1A | 5.3 | 6.0 | 16.4 | 12.7 | 2A | 8.5 | 13.1 | 6.6 | 8.4 | 3A | 11.4 | 21.2 | 48.3 | 40.0 |
| 1B | 1.8 | 2.3 | 19.4 | 23.5 | 2B | 7.2 | 15.2 | 6.4 | 11.5 | 3B | 4.9 | 3.1 | 36.0 | 30.0 |

TABLE 2

Mean Number of Trials Required to Attain Test Concept in First Study

| Group | Training Concepts | Test Concept | *N* | *N* Solving Test Concept | Mean Trials | *SD* | Range |
|---|---|---|---|---|---|---|---|
| 1A | Shape, size | Shape + size | 10 | 10 | 23.3 | 21.2 | 5–65 |
| 2A | Shape, color | Shape + size | 10 | 10 | 42.4 | 22.1 | 12–81 |
| 3A | Color, part-whole | Shape + size | 11 | 9 | 104.5 | 65.0 | 17–192 |

TABLE 3

MEAN NUMBER OF TRIALS REQUIRED TO ATTAIN TEST CONCEPT IN SECOND STUDY

| Group | Training Concepts | Test Concept | *N* | *N* Solving Test Concept | Mean Trials of *Ss* Solving Test Concept | Range |
|---|---|---|---|---|---|---|
| 1B | Shape, size | Color + part-whole | 10 | 1 | 167.0 | |
| 2B | Shape, color | Color + part-whole | 11 | 3 | 89.0 | 67–130 |
| 3B | Color, part-whole | Color + part-whole | 10 | 4 | 77.8 | 25–186 |

obtained; the group (1A) which had learned both of the simple concepts of which the test concept was composed learned the test concept more rapidly than did the group (2A) which had learned only one of the significant simple concepts, and this latter group learned it more rapidly than the group (3A) which learned neither of the appropriate simple concepts. The results of the second study failed to show the same *significant* differences, seemingly because such a large percentage of *S*s failed to acquire the test concept. The trend of the results are consistent, however, with the original hypothesis. It was found that the rate of learning of those *S*s who did solve the test concept appeared to be directly related to the number of *appropriate* simple concepts that had been learned. The data of the second study also suggested that the percentage of *S*s solving the test concept within the allotted trials was related to the number of appropriate simple concepts that had been learned.

There are several interrelated problems raised by the data, the implications of which are particularly relevant to the relationship between S–R theory and concept formation behavior. The first of these is the discrepancy between the results of Group 1A and 3B. It will be recalled that these two groups were treated similarly; they were initially trained on two simple concepts and then were required to learn a compound concept composed of the two simple concepts on which they had been trained. The results indicate that, whereas all *S*s in Group 1A were able to learn their compound concept, only 4 of the 10 *S*s in Group 3B were successful. An analysis of this discrepancy points to one of the basic mechanisms in human concept formation behavior.

The results of Group 3B demonstrate that "mastery" of the simple concepts does not necessarily insure the solution of the compound concept. If *S*s can learn the simple concepts, why cannot they learn the compound concept which is a mere combination of the two simple concepts? The answer to this question lies in the understanding of what responses are learned when *S* masters a simple concept. It should be remembered that when *S* was learning a simple concept, he was required to sort the response cards into only *two* categories. It would be possible for *S* to sort the cards appropriately without discovering the rationale for the correct sorting procedure. For example, in "solving" the part-whole concept, *S* by rote memory could learn that he should place all crescents and triangles below the large yellow crescent stimulus card and all squares and circles below the small black square stimulus card. Although such behavior results in the successful completion of the initial phase of the experiment, it fails to provide those verbal cues (part figures vs. whole figures) which greatly facilitate the acquisition of the test concept.

The essential implication of the above point is that the associations underlying the process of concept formation in a card-sorting test are not between the cards and the sorting response but rather between the cards and some verbal response to them which, in turn, serves as a cue for the sorting response. Our contention is that when *S*s of Group 1A learned the simple concepts of shape and

size, they were able to verbalize these conceptual differences, and these verbalizations served as cues which facilitated the acquisition of the test concept. Although it was probably easy for *S*s in Group 3B to verbalize the conceptual differences between the response cards when learning the simple concept of color, it was extremely difficult for these *S*s to conceptualize the differences between the response cards when sorted on the basis of the part-whole concept. It would seem that a triangle, especially the equilateral triangle which appeared on the cards used in this experiment, is unable, for most *S*s, to instigate the idea that it is part of a square.

The emphasis on the importance of verbal cues is not, of course, all speculative. The *S*s were questioned following the termination of the experimental session. As is so often the case, many *S*s were found to be amazingly inarticulate; this can undoubtedly be attributed in part to defects in our questions. Their replies, however, did suggest some interesting relationships. The *S*s in Group 1A were able to verbalize the rationale of their card sorting after having learned the test concept. Most of them were able to express themselves in terms of the abstract principles, i.e., the differences between the four categories was on the basis of the combination of both the size and shape factors. The remaining *S*s expressed themselves in a somewhat more concrete fashion, describing specifically the underlying features of their sorting procedure, e.g., "small figures with small figures, round figures with round figures, large figures with large figures, sharp-edge figures with sharp-edge figures." Contrasted with this were the verbal reports of the successful *S*s in Group 3B. Two *S*s were practically incoherent, being unable to spontaneously provide any rational basis for their correct sorting behavior. Upon further questioning, it appeared that these *S*s were sorting the cards correctly by responding on an abstract level to the differences in color but specifically to the differences in form; i.e., these *S*s learned to place achromatic crescents and triangles in one category, chromatic crescents and triangles in another, achromatic squares and circles in another, and chromatic squares and circles in a fourth category. Although they were able to conceptualize the differences between the achromatic and chromatic cards, they were unable to verbalize the differences between the cards containing part figures and whole figures. This type of solution, which is based upon a combination of both abstract and specific features was the most common one reported throughout the second study. Only one of the eight successful *S*s (*S* in Group 3A) was able to give a clear statement of the abstract principles which guided his sorting behavior.

The importance of verbal cues in concept formation is nicely illustrated, although not elaborated, by Oseas and Underwood (5). They used as stimuli three different geometrical forms of three sizes and shades. Size and shade were relevant dimensions, while the form was irrelevant. This resulted in a task involving nine concepts, each a different combination of size and shade. It was found that concepts based on small and large geometrical figures were learned more rapidly than those based on medium size figures. Oseas and Underwood attributed these results to interference produced by stimulus generalization.

> In learning concepts of medium size, generalizing tendencies . . . from both the small and large figures could interfere. But, in learning the concepts based on small figures, generalizing tendencies would be maximal from the medium-size figures, but minimal as regards the large because of their greater "distance" along the size dimension. Likewise, in learning the concepts based on large forms, interference again should be maximal from the medium-size figures but minimal from the small (5, p. 146).

The same line of reasoning should apply to concepts involving the three shades (white, gray, and black) along the brightness dimension. There was not, however, any difference in the speed of learning these concepts. The discrepancy between the learning of concepts based on size as compared with brightness would be due to the differences in the verbal responses produced by the stimuli according to our analysis. In the case of size, one does not initially respond

to the size dimension with the verbal responses of large, medium, and small. One has to learn that there are three sizes before these descriptive terms can be attached. It is during this learning of the existence of three sizes that the interfering effects of stimulus generalization are produced. In the case of shade, however, *S* has available from the very beginning three verbal responses (white, gray, and black) which differentiate the three points on the brightness dimension. Consequently, the interfering effects on the gray concepts arising from the stimulus generalization do not have any opportunity to operate. It would appear that the three distinctive words inhibit the expression of stimulus generalization based upon physical similarity, as was found to be the case in the study by Kuenne (3).

Returning to the results of our own study, it might be argued that our analysis of the discrepancy between the results of Groups 1A and 3B says no more than that *S*s in Group 1A were able to perceive the necessary relationships while most of the *S*s in Group 3B were unable to do this. We believe that such an argument is not entirely justified. Our emphasis has been on response-produced cues which serve to mediate formation of the appropriate card-sorting responses. Such an analysis points to what we believe are basic mechanisms as well as to the integration of thinking behavior and simple associative learning, rather than the creation of a hiatus between the two, as is customary by advocating a "perception-of-relationship" orientation. In addition to this, our analysis does not require that the appropriate relationships be perceived for the concept to be attained. For example, it would be our guess that the percentage of *S*s solving the test concept would increase by conditioning the triangle and crescent to a common nonsense syllable. Such a procedure would provide a common verbal cue which should facilitate the acquisition of the test concept, even though it would not emphasize the idea that a triangle is part of a square.

Let us now consider the significant differences found between the three groups in the first study. Our analysis would suggest that the advantage derived from the acquisition of the appropriate simple concepts did not stem from the mere repetition of correct sorting responses, but rather from the opportunity the initial training provided for the appropriate verbal responses to become dominant and thereby to facilitate the acquisition of the test concept. Essentially, the preliminary training, involving the appropriate simple concepts, modifies the cue situation rather than strengthening markedly any particular response tendency. Unlike a simple conditioning situation, in which the conditioning trials increase the habit strength between the conditioned stimulus and the conditioned response, the major effect of the training trials in the present study is to determine what verbal cues shall be dominant at the beginning of the test trials.

This point was borne out to some extent by *S*s' comments following the completion of their experimental session. The most common introspective report of *S*s in Group 1A, when describing the rationale of their correct sorting procedure for the test concept, was the simple statement of the abstract relationships involved (size and shape). The most common report of *S*s in Group 2A, who had only learned one (size) of the appropriate simple concepts, was a more specific description even though some abstract features were mentioned. An example of this type of report is one in which *S* categorized the groups as large squares and large triangles, large circles and large crescents, etc. The most common report of *S*s in the third group was vague; it failed to specify in any clear or simple manner the verbal cues which dictated *S*s' sorting response.

A basic problem raised by the present studies, but unfortunately not answered by them, is related to the question of whether those factors which influence the rapidity of acquisition of compound concepts (such as training on appropriate simple concepts) also influence the percentage of *S*s solving the compound concept. The data from our two studies would suggest that this is so since in the initial study the only two *S*s

who failed to solve the test concept were in the group that had no training on the appropriate simple concepts, while in the second study the percentage of solvers in each group was directly related to the number of appropriate simple concepts learned. One might argue, however, that the training on appropriate simple concepts has the effect of hastening the appearance of the appropriate verbal responses, but if *unlimited* time were afforded to all *S*s, the percentage of solvers would be unrelated to the type of preliminary training. Our findings give some support to this point since two *S*s in Group 1B (no training on appropriate simple concepts) who were unable to solve the test concept within the allotted trials were able to solve the test concept with additional trials. The theoretical problem involved in this question is whether the training on appropriate simple concepts merely hastens the appearance of the appropriate verbal response, assuming that it is within *S*s' repertoire, or whether this training "selects out" the appropriate verbal response which under other training conditions might not occur.

Our analysis of the present data, with its major emphasis on the cue situation being modified as a function of training, has certain similarities to Guthrie's (2) treatment of conditioning phenomena. This does not mean necessarily that only a simple contiguity conditioning theory, such as Guthrie's, is applicable to human thinking behavior. The stimulus situation is an important variable in all S–R theories, whether they be committed to a single-trial associative principle or to a habit-growth principle. Within a theoretical framework involving the habit-growth principle it can be said that the associations involved in the present study are not new associations with zero or relatively weak habit strengths, but rather they are established associations with their habit strengths approaching or having approached their asymptote values. The essential condition for the acquisition of the test concepts is that the relevant associations become dominant in their competition with the irrelevant associations. This comes about by rapid shifting in the dominant position in response hierarchy because of the almost equal strength of the competing associations involving the various verbal responses.

## Summary

An experiment was designed to determine the influence the learning of simple concepts (concepts based on one stimulus attribute) had upon the learning of a compound concept (concepts based on two stimulus attributes) which involved combinations of simple concepts. The experimental design consisted of two similar studies, each involving three groups, all of which were required initially to learn two simple concepts successively. Following this, each group was required to learn the test concept, which was a compound concept. One group had learned both of the simple concepts of which the test concept was composed, the second group had learned only one, while the third group had learned neither of these concepts. The results of the first study indicated that the rate of learning the test concept was directly related to the number of simple concepts appropriate to the test concept which had been learned. The trend of the results of the second study was consistent with that of the first study. It was found in the second study, however, that most of the *S*s, even those who had learned both of the appropriate simple concepts, failed to acquire the test concept.

Our analysis suggested that the advantage derived from the acquisition of the appropriate simple concepts did not stem from the mere repetition of correct sorting responses, but rather from the opportunity the initial training provided for the appropriate verbal responses to become dominant, and thereby facilitate the acquisition of the test concept.

## References

1. Berg, E. A. A simple objective technique for measuring flexibility in thinking. *J. gen. Psychol.*, 1948, **39**, 15–22.
2. Guthrie, E. R. *The psychology of learning.* New York: Harper, 1952.
3. Kuenne, M. R. Experimental investigation of the relation of language to transposition behavior in young children. *J. exp. Psychol.*, 1946, **36**, 471–490.
4. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
5. Oseas, L., & Underwood, B. J. Studies of distributed practice: V. Learning and retention of concepts. *J. exp. Psychol.*, 1952, **43**, 143–148.

(Received June 28, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# RATE OF RESPONSE DURING OPERANT DISCRIMINATION

MONCRIEFF H. SMITH, JR. AND WILLIAM J. HOY
*University of Washington*

The process of discrimination plays an important role in the analysis of behavior proposed by Skinner (6), but the method he used for studying it has several serious limitations. Skinner presented $S^D$ (the positive stimulus) and reinforced the first response made in its presence. He then withdrew $S^D$ and presented $S^\Delta$ (the negative stimulus) for a period of time during which no responses were reinforced, using the drop in rate of responding during $S^\Delta$ as the measure of discrimination formation.

This technique has at least two weaknesses. First, *E* cannot specify the exact nature of the stimulus that sets the occasion for the response. If $S^D$ be a light and $S^\Delta$ a dimmer light, the light intensity is only a part of the stimulus situation. The change from $S^\Delta$ to $S^D$ is a very important determiner of the response, as Skinner showed (6, p. 213) in his investigations of discrimination reversal. The possibility of a temporal discrimination is also apparent, since the time between reinforcements is constant. Since Skinner was not concerned with the discriminative capacities of the rats, these uncertainties did not disturb him.

The second difficulty is more serious. The method provides no good measure of the strength of the response in the presence of $S^D$. While $S^\Delta$ is present the response undergoes a change comparable to extinction, and this change may be observed in detail, in the rate of response. Since the rat continues to respond to the presence of $S^D$, it may be inferred that this response has not weakened, but its latency is the only measure of change in response strength.

The difficulty arises from the requirement that the response be reinforced in the presence of the positive stimulus. If the reinforcement is food, the time required to eat sets a limit on the rate of response, and if each of a series of responses is reinforced, rate of response becomes a meaningless measure. Several solutions to this difficulty have been proposed. Frick (3) used a situation in which $S^D$ and $S^\Delta$ were presented for alternate 2½-min. blocks of time, with every response during $S^D$ reinforced. The test of discrimination was made during a period in which $S^D$ was presented without reinforcement. This method allows a measure of response rate to $S^D$ and to generalized stimuli, but it prevents study of the formation of the discrimination.

Dinsmoor (2) used a technique he called periodic reinforcement of the response during $S^D$. His experiment consisted of a series of 3-min. cycles. For the first 2 min. of a cycle $S^\Delta$ was presented. $S^D$ was then presented for 1 min., and the last response reinforced. This allowed study of the formation of the discrimination, but made specification of the stimulus difficult. The discrimination was based partly on the stimulus controlled by *E*, and partly on the temporal interval, as shown by the finding that the rate of response began to increase toward the end of the 2-min. period of $S^\Delta$.

The method of discrimination training used in the experiment reported here seems to overcome the difficulties discussed above.

## METHOD

*Apparatus and discrimination procedure.*—The apparatus consisted of four identical Skinner boxes, which have been described by Kinney (5). The discriminative stimuli were two lights of different intensities. A hole 1 in. in diameter was cut in the wall of the box directly above the lever. This hole was covered with a pane of frosted glass, and a G.E. 7½-w. unfrosted bulb was mounted behind the glass (outside the box). The lamp could be operated on 75 v. (bright) or on 38 v. (dim). These stimuli are not specified exactly, since lowering the lamp voltage changed its color as well as its intensity, but the method could be used with carefully specified stimuli.

Discrimination training was given during a series of 2-min. trials, 20 trials to an experimental session. Either $S^D$ or $S^\Delta$ was present continuously during a 2-min. period. The sequence of $S^D$ and $S^\Delta$ was determined randomly, subject to the restrictions that each appeared ten times in a session and that neither appeared more than three times in a row. During one $S^D$ period three reinforcements were given; in three periods two were given; and in the remaining six periods, one was given. The order of these was randomly determined, as was the temporal point of reinforcement during the 2-min. period.

A control group was given the same sequence of stimulus changes and reinforcement, but reinforcement occurred half the time with one stimulus and half with the other, so that reinforcement was the only possible discriminative stimulus. Analysis of control group data into the ten periods with, and the ten without, reinforcement provided a measure of responses gained or lost due to the administration of reinforcement.

Between 2-min. periods *E* turned off the main power switch, which disconnected the discriminative stimuli, the overhead light in the boxes, and all control and recording apparatus. Stimulus conditions were set manually for the next period, which took about 30 sec. The *E* set up reinforcements manually, following a prearranged schedule that was made up independently for each rat on each day. After the reinforcement was set, the next bar depression by the rat released it. At the end of the 2-min. period all "unclaimed" reinforcements were canceled.

It is felt that this randomizing of all factors except the correlation between reinforcement and light intensity makes it possible to eliminate all possible sources of discrimination except one. In addition, this technique allows continual observation of the rate of response in the presence of both positive and negative stimuli.

*Subjects.*—The *S*s were 24 albino rats, Wistar strain, about 100 days old at the beginning of the experiment. After a preliminary period of handling, 28 rats were trained to respond to the sound of the feeder in the boxes and allowed 30 reinforced responses. They were then run for one day (20 periods) under control group conditions (discriminative stimuli present but uncorrelated with reinforcement). On the basis of this performance three matched groups of eight rats each were formed, and the remaining four discarded.

*Experimental procedure.*—One group of eight *S*s had the bright light positive; another dim light positive; and the third was the control. Throughout the experiment the rats were run in sets of four. Each set contained at least one representative of each of the three groups. Discrimination training began after the groups were matched and was continued for 27 days. Without interruption discriminative stimuli were reversed for the two experimental groups, and all rats were run for an additional 33 days. Five days of extinction with discriminative stimuli present followed this.

At the end of each experimental session *S*s were given food for 1 hr. in their home cages, so that they were deprived of food for 22 hr. at the beginning of the next session. A reinforcement consisted of ½-gm. powdered lab chow, and each *S* got a maximum of 15 of these in a session.

At the end of Day 25 of the original discrimination, *S*s were not fed, so that on Day 26 deprivation conditions were increased.

One of the control rats became sick on the third day of extinction, so he was not included in the extinction data. He was included, however, for the 60 days of discrimination training.

## RESULTS

Eight of the experimental *S*s started out with the bright light positive, and eight with dim positive. Examination of the discrimination data and of total rate of response showed no differences between these groups. Moreover, an analysis of a six-day sample of data from the control group showed that response rate did not differ for bright and dim lights. Since the difference in light intensities used was not sufficient to evoke a differential aversion to the light, the two experimental groups are combined for most comparisons.

Our primary concern was with the effect of discrimination on over-all response rate (algebraic sum of response rates during $S^D$ and $S^\Delta$). Figure 1 gives a partial answer to this

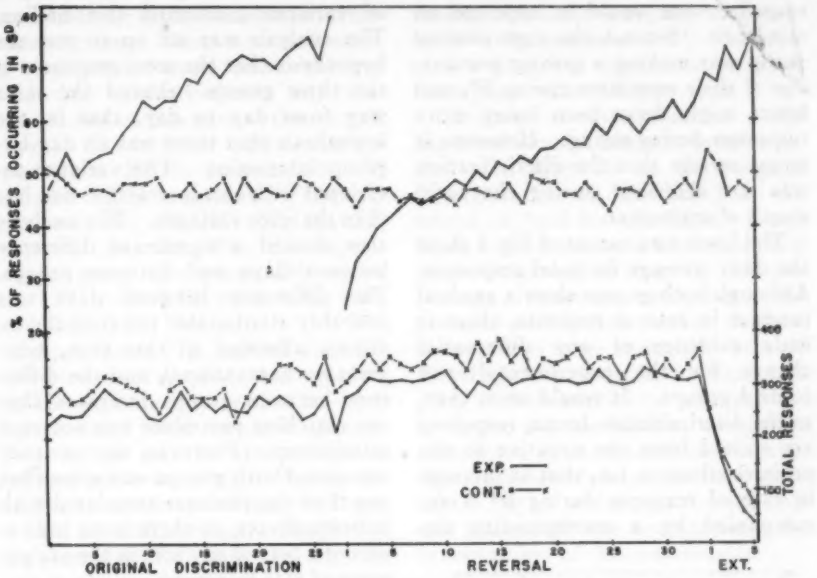


FIG. 1. Average daily performances for discrimination and control groups during acquisition, reversal, and extinction of a brightness discrimination. The top two lines show the daily average of percentage of total responses made in the presence of the positive stimulus. The lower two lines show the daily average number of responses made in an experimental session (in the presence of positive or negative stimuli).

question. The two upper curves show the formation of the original and the reversed discrimination, in the form of daily averages of percentage of total responses made in the presence of $S^D$. This percentage was computed for each *S*, and the percentages averaged to make the plot. The absence of any systematic shift in the percentage figure for the control group indicates that the reinforcement itself did not come to serve as a discriminative stimulus. The fact that the control group is consistently below 50% indicates that the responses lost during eating were not made up after eating.

During extinction the designation of $S^D$ periods for the control group was completely arbitrary, and there were no possible differential cues for the rats. Since there were no more reinforcements, there were no more responses lost due to eating and the control group percentage would be expected to rise to 50%. Any deviations from this figure must be attributed to chance. The experimental group also showed an increase in percentage, which was probably due in part to the elimination of eating; but it was a greater rise than was shown by the control group.

The increases of the two groups during extinction were compared by a *t* test. This comparison was based on the individual increases for the average percentages from the last three days of discrimination to the five days of extinction. The average increase was 8.4 for the experimental group ($N = 16$) and 1.7 for the control ($N = 7$), with $t = 2.51$ and $p = .05$. Two factors reduce the validity of this comparison. First, the experimental group was still climb-

ing before extinction started, so that some increase would be expected on this basis. Second, the experimental group was making a greater percentage of their responses during $S^D$, and hence might have been losing more responses during eating. However, it seems certain that the discrimination was not abolished during the early stages of extinction.

The lower two curves of Fig. 1 show the daily average for total responses. Although both groups show a gradual increase in rate of response, there is little evidence of any differential change for the experimental and control groups. It would seem that, as the discrimination forms, responses are shifted from the negative to the positive stimulus, i.e., that an increase in rate of response during $S^D$ is accompanied by a corresponding decrease in rate during $S^\Delta$. An analysis of variance confirmed this finding. The analysis was set up to test the hypothesis that the total responses of the three groups behaved the same way from day to day, that is, the hypothesis that there was no day-by-group interaction. The variance associated with this interaction was less than the error variance. The analysis also showed a significant difference between days and between groups. The difference between days was probably attributable to external conditions affecting all rats (e.g., temperature fluctuations), and the difference between groups indicates that our matching procedure was not very satisfactory. However, the variance associated with groups was somewhat less than the variance associated with individual rats, so there is no indication of a bias of any sort in the assignment of rats to groups.

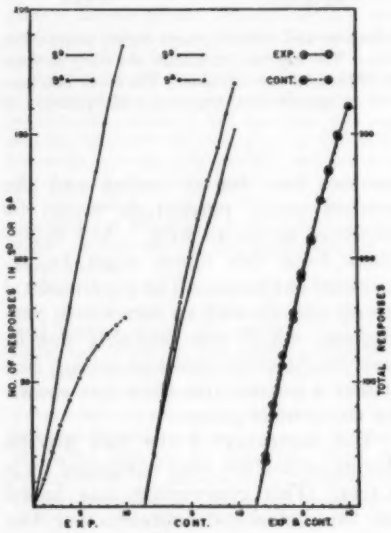


FIG. 2. Within-session changes in response rate, based on the data from Days 21 through 25 of original discrimination. The first two sets of curves compare cumulative total number of responses in the presence of the positive and negative stimulus, for the experimental and control groups. The third pair of curves compares the form of the curve for total responses within a session for the experimental and control groups.

Further evidence on the relationship between discrimination and over-all response rate is given in Fig. 2. During each daily period, the effectiveness of discrimination increased, as shown by the increasing separation of the first two curves. These curves were obtained by cumulating the responses made in the presence of $S^D$ and $S^\Delta$ within the daily session. The sequence of positive and negative stimuli was determined independently for each *S* on each day, so that the *n*th presentation of $S^D$ did not occur at the same time in the session for all *S*s. These differences were ignored, and only the sequence of positive (and negative) stimuli was preserved. Figure 2 is based on the average of all *S*s in a group over a period of five days, Days 21 through 25 of the original discrimination. The percentage figures of Fig. 1 indicate that the extent of discrimination was increasing somewhat during this period, but the general form of the curves in the first part of Fig. 2 was typical of all *S*s on all five days.

The second pair of lines in Fig. 2 represents the same cumulation for the control group. There is no indication of a systematic trend within the session that might indicate any discrimination on the part of the control group. The ratio of cumulative responses in $S^D$ to responses in $S^\Delta$ remains fairly constant. (The sequence is .96, .91, .89, .88, .88, .87, .87, .88, .88, .88. For the experimental group the comparable ratios are 1.76, 1.73, 1.78, 1.84, 1.90, 2.01, 2.11, 2.22, 2.34, 2.44.)

Inspection of Fig. 2 shows that the over-all rate of response declines during the experimental session for both experimental and control groups. This decline may be attributable to reduced motivation, since during the session each *S* received almost 2 gm. of food as reinforcement. Because of the pronounced drop in the response rate in the presence of $S^\Delta$, it appears that the experimental group rate drops more during the session than does the control rate. The third pair of curves in Fig. 2 shows this is not the case. To get these curves, the average cumulated total responses for $S^D$ and $S^\Delta$ were added together, to give an approximation of the form of the average cumulative curve for all responses. The control group consistently gave a higher total number of responses, so that comparison of the forms of the curves was facilitated by multiplying all values for the experimental group by a constant chosen to bring the curves together at their end point. This comparison shows that the control group actually had a greater drop in rate during the session than did the experimental group. This last finding was not verified by the records of the other days, and there is considerable support for the statement that the form of the cumulative response curve within the day is the same for the experimental and control groups. Thus, although the discrimination builds up within each daily session, the discrimination is not accomplished by dropping responses made in the presence of $S^\Delta$. There is such a drop, but compensation for it takes place while $S^D$ is present.

The test of discrimination under increased deprivation was uninformative. On Day 26 of the original discrimination all rats had been deprived of food for two days. There was a slight increase in over-all rate of responding, and a drop in the percentage of responses occurring in the presence of $S^D$, but neither of these changes was large enough to be very revealing. In general the data suggest, both in within-day changes and in day-to-day variability, a negative correlation between over-all response rate and effectiveness of discrimination. Better techniques for the control of over-all rate are necessary for investigation of this relationship.

## Discussion

Skinner (6) treats discrimination learning in terms of a separation in strength of two related reflexes. He has dealt primarily with the situation in which the two reflexes share a common response, and in this case he assumes that the operant in the presence of the negative stimulus undergoes extinction from failure of reinforcement, while the same operant in the presence of the positive stimulus is maintained, or increased, in strength. The extinction of the "incorrect" reflex is slowed by a process of generalization, or induction.

Some of Skinner's statements can be interpreted to mean that in a discrimination situation such as ours the strength of the operant (the bar press) will remain unchanged during discrimination, and that it will be equal to the sum of the strengths of the two reflexes. However, little or no experimental data have been offered to support this interpretation, and it could well be that the response in the presence of the positive stimulus remains unchanged, while the response in the presence of the negative stimulus undergoes extinction. Our data show that over-all response rate does remain

unchanged during the acquisition of a discrimination and its reversal, and thus support the first interpretation.

However, these data suggest that some of Skinner's conclusions about induction may have been based on artifacts of the discrimination situation he used. Discrimination reversal in our situation was a very difficult task. Comparable rises in discrimination effectiveness took longer for the reversal than for the original discrimination, and there is no reason to believe that a second reversal would have been any less difficult. In Skinner's situation the first reversal is easy, and the second almost instantaneous, supporting his conclusion that induction breaks down during discrimination. Although Skinner recognized that any change in stimulus conditions (i.e., from bright to dim or from dim to bright) could become a discriminative stimulus, it would seem that he underestimated the importance of this aspect of his stimulus situation. We suggest that his experimental technique does not allow a true reversal of a discrimination.

Another interesting aspect of the present results is the increase in discrimination during extinction. It is clear that we are dealing here with the extinction of a discriminative response, rather than the abolition of a discrimination (by reinforcing both responses). Keller and Schoenfeld (4) make this distinction clearly, but it tends to get overlooked in most discussions of discrimination—perhaps because instances of pure extinction rarely occur. The usual result when a discriminative response is no longer reinforced with food is that some response continues, but selectivity breaks down. For example, Deese (1) found that during the extinction of a T-maze performance the rats continued to run, but ceased to select the "correct response." One wonders if this situation is not one in which uncontrolled reinforcement continues to operate, following either response. If this is the case, it might be better to recognize this as abolition of a discrimination, and to supply controlled reinforcement for both responses.

Of course, when response tendency drops to a very low level, evidence of discrimination tends to disappear. Some of our rats had practically stopped responding after five days of extinction, and their percentages were becoming too variable to mean much. There is no reason to believe, however, that the two responses had become equal in strength. For example, it is doubtful that after complete extinction the rats could learn a reversed discrimination as easily as they could relearn the old one.

## Summary

A new technique was applied to the study of operant discrimination in the Skinner box. The experimental session was divided into 20 periods of 2 min. each, with $S^D$ present during ten periods, and $S^\Delta$ for the other ten. Responses were reinforced aperiodically during the positive periods.

Sixteen rats were run under these conditions for 27 days original discrimination; then 33 days on the reverse discrimination; then 5 days extinction. A control group of eight rats was given the same sequence of stimuli and reinforcement, but reinforcement was not correlated with either stimulus. The results were:

1. As the discrimination formed, the over-all rate of responding remained constant, and the responses shifted from the negative periods to the positive ones. This held true during the learning of the reversed discrimination, too.

2. During extinction the discrimination was not abolished. Response rate dropped in the presence of both stimuli, but relatively more rapidly in the presence of the negative stimulus.

## References

1. Deese, J. The extinction of a discrimination without performance of a choice response. *J. comp. physiol. Psychol.*, 1951, **44**, 362–366.
2. Dinsmoor, J. A. The effect of periodic reinforcement of bar-pressing in the presence of a discriminative stimulus. *J. comp. physiol. Psychol.*, 1951, **44**, 354–361.
3. Frick, F. C. An analysis of operant discrimination. *J. Psychol.*, 1948, **26**, 93–123.
4. Keller, F. S., & Schoenfeld, W. N. *Principles of psychology*. New York: Appleton-Century-Crofts, 1950.
5. Kinney, G. C. The effects of adaptation on operant responses. Unpublished master's thesis, Univ. of Washington, 1953.
6. Skinner, B. F. *The behavior of organisms*. New York: D. Appleton-Century, 1938.

(Received February 1, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# PERFORMANCE IN SELECTIVE LEARNING AS A FUNCTION OF HUNGER[1]

CHARLES K. RAMOND[2]

*State University of Iowa*

In selective learning situations, in which the relative frequency of occurrence of two or more mutually incompatible responses is measured, conflicting results have been obtained with regard to the role of the hunger drive ($D$), defined in terms of hours of food deprivation ($T_D$). Some studies (**8, 15, 16, 17**) have found choice behavior ($R\%$) to be a positive function of $D$, while other studies (**5, 13, 14**) have found no such relationship.

The above results collectively imply no unique theoretical role of the hunger drive in selective learning, although most of them are consistent with the familiar Hullian notion (**1**) that drive strength ($D$) multiplies a learning factor, habit strength ($H$). Nevertheless, in many of these same experiments, quite different results could have been obtained without having provided a basis for the rejection—or conclusive proof—of the $D \times H$ hypothesis.

One such experiment, by Loess (**5**), discussed the above barrier to an evaluation of the $D \times H$ hypothesis, and proposed a method of eliminating it. Loess's apparatus, which was also used in the present experiment (see Fig. 1), consisted of a double-bar Skinner-type box in which either or both of two bars could be presented. On the first trial of each block of trials, both bars were presented, while on the subsequent trials of each block (called forced trials), a single bar only was present. Choice data from the free trials provided a measure of performance in selective learning; latency data from the forced trials provided a measure of performance in simple instrumental learning. All animals were induced to choose one bar rather than the other by manipulation of delay of reinforcement ($K$): a response to one bar was followed by reinforcement after 1 sec., whereas a response to the other bar was reinforced after 5 sec. As a consequence, animals came in the course of training to choose the short-delay bar on free trials, and, on forced trials, to respond to the short-delay bar more quickly than to the long-delay bar.

Loess found that a 22-hr. hunger drive group did not choose the short-delay bar significantly more often than did a 4-hr. hunger drive group. He pointed out, however, that this finding may or may not support the $D \times H$ hypothesis, depending upon the assumption made concerning the interaction between $D$ and $K$.[3] On the other hand, Loess noted

[1] The present paper is a portion of a dissertation submitted to the Graduate College of the State University of Iowa in partial fulfillment of the requirements for the Ph.D. degree. The writer is indebted to Professor Kenneth W. Spence, who suggested the problem, and to Dr. W. R. McAllister, who supervised the collection of the data. The drawing of the apparatus in Fig. 1 is reproduced through the courtesy of Dr. H. B. Loess.

[2] 1st Lt., U. S. Army. Now Acting Military Chief, Human Research Unit No. 3, Office, Chief of Army Field Forces, Fort Benning, Georgia.

[3] Hull (**1**) has assumed a multiplicative interaction between $D$ and $K$ while Spence (**11**) has suggested that $D$ and $K$ combine according to simple arithmetical addition. The deductions from these two assumptions are, respectively, that $R\%$ in the differential-delay situation (*a*) is and (*b*) is not a function of $D$. The reader may verify this by substituting $(D \times K)$ and $(D + K)$ for $X$ in the equation

$$R\% = f(XH_A - XH_B).$$

Recent evidence (**10**) favors the additive assumption, at least for the values of $D$ used by Loess. If this is the case, then Loess's findings support the $D \times H$ hypothesis.

that the effect of drive on selective learning *can* be unequivocally deduced—i.e., without making further assumptions about the combination of $D$ and other motivational factors (e.g., $K$ or $V$)—for a situation in which no variables other than $D$ and $H$ are manipulated. In such a situation, choice is induced by varying the determinant of $H$, number of reinforced trials ($N$). Each $S$ receives, between successive choice trials, more forced trials to one bar (Bar A) than to the other (Bar B). The deduced prediction concerning the choice behavior of differentially deprived groups in this situation is that the high-drive group should choose the more frequently reinforced bar (Bar A) more often than should the low-drive group. The assumptions necessary for this deduction may be specifically stated as follows:

1. The $D \times H$ hypothesis: Hull (1) has assumed that excitatory potential ($_sE_R$), as reflected by such measures of response strength as response speed, is, other factors being constant, a positive function of a multiplicative combination of drive strength ($D$) and habit strength ($H$), a relationship mathematically expressible by the equation,

$$_sE_R = f(D \times H).$$

2. A relationship between choice behavior and differential excitatory tendency: from Hullian definitions and postulates, Logan (6, 7) has derived, and empirically tested in the differential-delay selective learning situation, an equation which states that choice behavior ($R\%$) is a positive function of the difference between the excitatory potentials of the responses between which choice is made, i.e.,

$$R\% = f(_sE_{R_A} - {}_sE_{R_B}).$$

From these two assumptions choice behavior in the situation under consideration may be deduced as follows:

$$\begin{aligned} R\% &= f(_sE_{R_A} - {}_sE_{R_B}) \\ &= f(DH_A - DH_B) \\ &= f[(D)(H_A - H_B)]. \end{aligned}$$

Thus choice behavior in this experiment was predicted to be a positive function of drive, i.e., a high-drive group should choose Bar A more often than should a low-drive group. The present experiment tested this prediction in order to assess the tenability of the assumptions from which it was derived.

## Method

*Subjects.*—Forty-eight hooded rats from the colony maintained by the department of psychology of the State University of Iowa served as *Ss*. All rats were naive and between 95 and 120 days old at the beginning of the experiment. The *Ss* were assigned to two drive groups: 24 *Ss* always received training 22 hr. after feeding and 24 received training 4 hr. after feeding. This assignment was not random; effort was made to have an equal number of rats from each litter in each drive group. The consequence of this litter-matching technique for the subsequent analysis of variance is discussed below.

Sixteen female rats constituted an initial replication, while 32 male rats were later run in a second replication. Since sex proved to be insignificant both as a main effect and as an interaction component, the data from the two replications were pooled.

*Apparatus.*—The apparatus, a double-bar Skinner-type box, is shown in Fig. 1. It has been described in detail elsewhere (5, 6). Briefly, it consisted of a triangular response chamber 18 in. across at the rear, 12 in. deep, and 10 in. high. A starting box led into the front apex through an outer brass door and an inner door of glass. Either or both of two bars

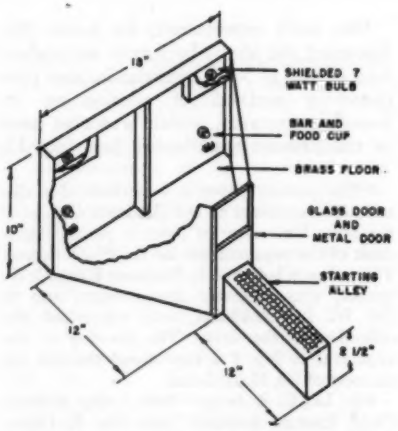


Fig. 1. A cutaway view of the double-bar Skinner-type box

could be presented through holes placed near the edges of the wall opposite the doors.

Raising the glass door started a Standard Electric Timer. By touching a bar, the rat completed a low-ampere contact relay circuit (2) between the brass floor of the box and the bar. Closure of this circuit stopped the timer, retracted the bar(s), and fired a Hunter Decade Interval Timer (3). One second later an .08-gm. pellet of food dropped into a food cup situated just below where the bar had been. A 1000-cycle tone sounded concurrently with the release of the pellet and was turned off manually after *S* obtained the food. The apparatus was in a dark, sound-resistant room, and was painted flat black, except for the floor of the response chamber, which was unpainted brass plate. A shielded 7-w. bulb was located above each food cup; the total illumination level in the response chamber was such that the food cup could not be seen by *E* (nor, presumably, by *S*) unless the bar light above it was on.

*Feeding schedule.*—All *S*s were placed on a 22-hr. feeding schedule for seven days prior to preliminary training during which time they were fed 8 gm. of dry Purina chow in individual feeding cages at the same hour every day.

*Preliminary training.*—Beginning on the eighth day, each rat received 1 hr. of preliminary training daily for three days in order to eliminate irrelevant exploratory behavior during learning, and to establish the secondary reinforcing properties of the 1000-cycle tone. The bars were never present during these preliminary training periods. Each rat was placed in the response chamber with one bar light on and a pellet of mash in the food cup beneath. Thirty seconds after *S* obtained this food and was facing in the direction of the food cup, another pellet was delivered, accompanied by the sound of the 1000-cycle tone. The tone was turned off when *S* appeared to have eaten the pellet. In this manner, four more pellets were given at approximately 30-sec. intervals, whereupon *S* was removed from the apparatus for 30 min. and then given similar training at the opposite food cup. Half the *S*s received their first training at the left food cup and the other half at the right food cup. Preliminary training thus consisted of the receipt of 30 pellets of food, 10 per day (5 at each cup), each accompanied by the sound of the 1000-cycle tone.

*Experimental training procedure.*—Training proper began the day after preliminary training and continued every day thereafter for 22 days. On each trial, *S* was placed in the starting alley, and the metal door was raised as soon as *S* faced the response chamber. Three seconds later the glass door was raised, permitting entrance to the response chamber.

On free trials, both bars were presented with

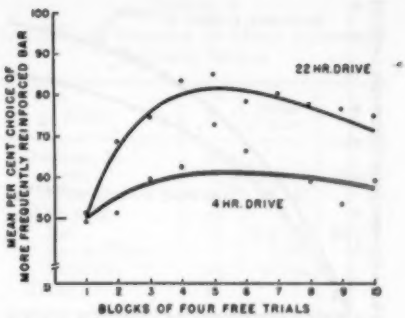


FIG. 2. Per cent choice of the more frequently reinforced bar (Bar A) as a function of drive

both bar lights lighted; on forced trials only one bar was presented and illuminated. Trials were administered in blocks of three, the first trial being free and the last two forced. The forced trials were arranged so that *S* received, in any block, two trials to one bar and one to the other. If A represents the more frequently reinforced bar and B the less frequently reinforced bar, the pattern of forced trials may be described as follows: if *S* chose Bar B on its free trial, it was then forced twice to Bar A on the next two trials. But if *S* chose Bar A on its free trial, it was then half the time forced to A, then B, and half the time forced to B, then A. Thus the pattern of the trial blocks was always one of the following: BAA, AAB, or ABA.

The consequence of this procedure was that immediately prior to any free-choice trial *S* had received twice as many reinforced trials to one bar as to the other. For a random half of the rats, Bar A was on the left; for the other half it was on the right. Each animal was given 120 trials, 80 to Bar A and 40 to Bar B. These trials were administered at the rate of one block of three trials per day for the first four days; thereafter two blocks (six trials) were run per day. Latencies were recorded to the nearest hundredth of a second on all trials, free and forced.

## RESULTS

The choice behavior of the 22-hr. and 4-hr. drive groups on free trials is shown in Fig. 2 by the solid and open circles, respectively. The corresponding smooth curves were derived from the latency data on forced-choice trials by a method described elsewhere (7) and discussed below. Three aspects of these choice curves

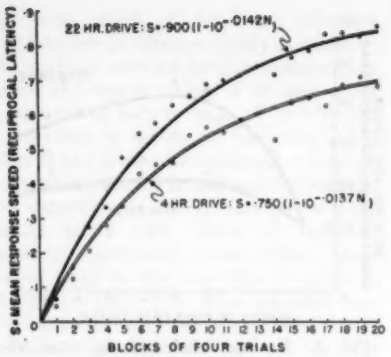


FIG. 3. Response speed on forced-choice trials to the more frequently reinforced bar (Bar A) as a function of drive

may be noted. First, both groups developed a temporarily significant preference for the more frequently reinforced bar. For an $N$ of 24, 67% differs significantly ($p = .05$, one-tailed test) from the chance level of 50%; it will be seen that the 22-hr. group exceeded this value on its last eight trial blocks, while the 4-hr. group surpassed 67% on only one block of trials.[4]

Secondly, the high-drive group chose the more frequently reinforced bar more often than did the low-drive group. To evaluate this difference statistically, each $S$'s total number of choices of Bar A over the 40 free trials was tallied, and the resulting distributions compared for the two drive groups. Since these distributions were markedly skewed, the nonparametric Mann-Whitney $U$ test (9) was employed. The obtained value of z was 2.49, which permitted rejection of the null hypothesis at beyond the 1% level of confidence (one-tailed test).

Thirdly, these choice curves differed markedly in form from typical selective learning curves in that they reached a maximum and then slowly declined. This downward trend was reliable in that the combined groups chose Bar A significantly more often during the fifth block of free-choice trials than during the tenth block ($z = 1.74$, $p < .05$, one-tailed test). The reversal of the choice curve was not unexpected, and the approximate trial on which maximum choice would occur was predicted before the experiment began. The shape of the curves is discussed below.

Figure 3 shows the performance of the two groups on forced trials to the more frequently reinforced bar (Bar A). The smooth curves are typical Hullian exponential functions, the parameters of which were obtained by fitting straight lines to plots of the logarithms of response speed against $N$ by the least-squares method.[5] The response measure of group mean speed for any four-trial block was determined by averaging subject means of those reciprocal latencies during the four trials which were to Bar A. Data for the 40 trials to Bar B, not shown here, were similar to the first half of the Bar A data, and were adequately fitted, as would be expected, by the Bar A smooth curves.

A mean "asymptotic" response speed to each bar was computed for each $S$ by averaging its speeds on all

[4] A greater maximum preference for Bar A, and thus a wider range of per cent choice values, would have increased the reliability of the observed relationship between choice and differential speed. Preference for Bar A could be increased by using an A:B trial ratio higher than the present 2:1, say 5:2 or 3:1.

[5] It may be of interest to note the stability of the $i$-parameter in equations of the form, response speed $= a(1-10^{-iN})$, as they have been fitted to various data obtained in the present apparatus. Exponential curves fitted by the writer to speed data collected by Logan (6) and Loess (5) yielded $i$-parameters of .0146 and .0118, respectively. Stein (12) has obtained an $i$-parameter of .0108 from a group similar, except for strain, to the 22-hr. drive group of the present study.

forced trials in the last 20. Group means of these speeds are given in Table 1. The measures were analyzed in a Type I (4) or "split-plot" design, and both bar and drive differences were found to be highly significant (see Table 2).

A statistical "error" typical of many designs is present in this analysis in that the variance of a factor (litter) was removed experimentally, by counterbalancing, but not statistically, by subtracting a sum of squares. On the assumption that litter does not interact with drive, a litter sum of squares could have been computed and removed from the "between" error sum of squares. This was unnecessary since the main effect of drive was already clearly significant, even with the spuriously inflated error term.

## Discussion

In accordance with the considerations presented in the introduction, the significantly superior choice behavior of the 22-hr. group may be interpreted as supporting the two assumptions from which this result was derived, namely, the $D \times H$ hypothesis and the positive relationship between choice and differential speed.

Specifically, this latter relationship, as derived by Logan (7), equates the normal deviate ($X_C$) corresponding to per cent choice to a linear function of the algebraic difference between response speeds on immediately prior forced trials, i.e., $X_C = a(\text{Speed}_A - \text{Speed}_B) + b$, where $a$ and $b$ are constants. Since $R\%$ is by definition ogivally related to $X_C$ and since speed may be related to $N$ by exponential equations such as those in Fig. 3, it is apparent that an equation of the type $R\% = f(N)$ derives directly from Logan's linear function. The smooth curves of Fig. 2 are two plots of this equation, having been derived from the speed data of forced trials by the procedure described above. Practical considerations preclude a presentation here of the derivation of this equation; however, some appreciation of the general nature of the relationship between choice and differential speed may be had from the following discussion of the unusual shape of the derived per cent choice curves.

If per cent choice of one bar is a positive function of the difference between the speeds to the respective bars, i.e., if $R\% = f(\text{Speed}_A - \text{Speed}_B)$, then the larger or smaller the difference between the speeds, the greater or lesser will be the per cent choice of the more frequently reinforced bar. The maximum of the per cent choice curve, by this reasoning, corresponds to the point of maximum difference between the speeds to the two bars. Furthermore, once the faster rising curve for Bar A reaches its asymptote, the difference between the speeds to the two bars will begin to diminish, since the slower growing curve for Bar B is still rising. This lessening of the difference between the speeds to the two bars is reflected in the decline of the per cent choice curve. If enough trials were run to allow the slower growing Bar B curve to reach its asymptote (which should, of course, be the same as

TABLE 1

Group Mean Asymptotic Speeds

| Drive Group | Bar A | Bar B |
|---|---|---|
| 22 hr. | .818 | .610 |
| 4 hr. | .678 | .508 |

TABLE 2

Analysis of Variance of Mean Asymptotic Response Speeds

| Source | *df* | Mean Square | *F* |
|---|---|---|---|
| Between *Ss* | 47 | | |
| Drive | 1 | 352,959 | 18.83* |
| Error (*b*) | 46 | 18,745 | |
| Within *Ss* | 48 | | |
| Bar | 1 | 858,249 | 84.91* |
| Bar × Drive | 1 | 6,684 | .66 |
| Error (*w*) | 46 | 10,108 | |
| Total | 95 | | |

* Significant beyond the .001 level.

the asymptote for Bar A), then the per cent choice curve should, theoretically, slowly return to 50%. This reversal downward is apparent in Fig. 2 and was, as stated above, statistically significant.

It will be noted that the shape of the *R*% curves also constitutes evidence for the asymptote of habit strength. The negative acceleration which characterizes performance curves in this and many other instrumental learning situations may conceivably be explained in terms of shrinkage of the unit of response measure as the physiological limit of performance is approached. But if the above discussed relationship between choice and differential speed is granted, then the fact of the choice-curve reversal precludes an explanation in terms of unequal response units, hence suggests that the learning curve is indeed negatively accelerated and approaches a real asymptote.

## Summary

Forty-eight rats, divided equally into 22-hr. and 4-hr. hunger drive groups, learned to choose one bar rather than another during 40 free trials in a double-bar Skinner-type box. This choice behavior was induced by the administration of 80 interspersed forced trials to the individual bars such that immediately prior to any free trial, a rat had received twice as many reinforced trials to one bar as to the other.

The high-drive group chose the more frequently reinforced bar significantly more often than did the low-drive group. The high-drive group also ran significantly faster to both bars on forced trials than did the low-drive group.

The difference in choice behavior between the drive groups was interpreted as lending support to Hull's assumption that response strength is a positive function of a multiplicative combination of drive strength and habit strength.

The shape of the per cent choice curves was interpreted as supporting the notion that choice behavior may be meaningfully related to the difference between the strengths of the incompatible responses between which choice was made.

## References

1. Hull, C. L. *Principles of behavior*. New York: D. Appleton-Century, 1943.
2. Hunter, T. A. An electronic contact relay. *Amer. J. Psychol.*, 1952, **65**, 458–459.
3. Hunter, T. A., & Brown, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
4. Lindquist, E. F. *Design and analysis of experiments in psychology and education*. Boston: Houghton Mifflin, 1953.
5. Loess, H. B. The effect of variation of motivational level and changes in motivational level on performance in learning. Unpublished Ph.D. dissertation, State Univ. of Iowa, 1952.
6. Logan, F. A. The role of delay of reinforcement in determining reaction potential. *J. exp. Psychol.*, 1952, **43**, 393–399.
7. Logan, F. A. Three estimates of differential excitatory tendency. *Psychol. Rev.*, 1952, **59**, 300–307.
8. MacDuff, M. M. The effect on retention of varying degrees of motivation during learning in rats. *J. comp. physiol. Psychol.*, 1946, **39**, 207–240.
9. Mann, H. B., & Whitney, D. R. On a test of whether one of two random variables is stochastically larger than the other. *Ann. math. Statist.*, 1947, **18**, 50–60.
10. Ramond, C. K. Performance in instrumental learning as a joint function of delay of reinforcement and time of deprivation. *J. exp. Psychol.*, 1954, **47**, 248–250.
11. Spence, K. W. Current interpretations of learning data and some recent developments in S-R theory. In *Learning Theory, Personality Theory and Clinical Research*. New York: Wiley, 1954.
12. Stein, L. Performance in selective learning as a function of delay of reinforcement. Unpublished M.A. thesis, State Univ. of Iowa, 1953.
13. Teel, K. S. Habit strength as a function of motivation during learning. *J. comp. physiol. Psychol.*, 1952, **45**, 188–191.
14. Teel, K. S., & Webb, W. B. Response evocation on satiated trials in a T-maze. *J. exp. Psychol.*, 1951, **41**, 148–152.
15. Tolman, E. C., & Gleitman, H. Studies in spatial learning: VII. Place and response learning under different degrees of motivation. *J. exp. Psychol.*, 1949, **39**, 653–659.
16. Tolman, E. C., & Honzik, C. H. Degrees of hunger, reward, and non-reward, and maze-learning in rats. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 241–256.
17. Tolman, E. C., Honzik, C. H., & Robinson, E. W. The effect of degrees of hunger upon the order of elimination of long and short blinds. *Univ. Calif. Publ. Psychol.*, 1930, **4**, 189–202.

(Received February 11, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# EXTINCTION BY OMISSION OF FOOD AS A FUNCTION OF GOAL-BOX CONFINEMENT[1]

WALTER C. STANLEY AND MARC I. ROWE

*Brown University*

It is typical extinction procedure with runways, alleys, and mazes for *E* to confine *S* in the goal box or on a goal platform for about the same amount of time as was allowed during training. During extinction, however, food is absent and the food-deprived *S* is exposed to stimuli previously contiguous with eating, although eating food is now impossible.[2] When so confined, the typical rat is quite active. Although the component responses of this activity have not been studied in their own right, it will be granted that initially this activity consists of responses such as scratching, sniffing and biting, e.g., a closed *food* compartment door. If *S* remains confined in the goal box, other responses appear; e.g., locomotion, face washing, jumping, climbing, scratching and biting the *goal* box door, defecating, urinating, and sometimes "freezing."

Since the latter responses occur, contiguity theory (**1, 2, 3**) assumes that each, in turn, is conditioned to that fraction of contiguous stimuli not already conditionally related to the occurring response; thus, all such contiguous stimuli, in turn, lose their conditional relation to any incompatible responses. Assuming that moving forward in the goal box is incompatible with many of these "new" responses, moving forward should lose its conditional relation to whatever stimuli are present whenever "new" responses occur.

The "extinction" of moving forward in the goal box, or, stated positively, the "acquisition" of any responses incompatible with moving forward in the goal box should lead to loss of the entire response sequence of running from start box to goal box. The progressive elimination of the running sequence should take place, as trials continue, because competing responses should be conditioned to stimuli present earlier and earlier in the sequence. More sudden disruptions of the running sequence could occur if competing responses generalize to any part of the apparatus similar to the goal box, or to *S*'s being released in the start box by *E*.

A long goal-box confinement should provide more opportunity for the occurrence of competing responses than a short goal-box confinement. Therefore, a long goal-box confinement should produce faster extinction of the running sequence than a short confinement. The experiment to be described tested this expectation with 10-, 30-, and 60-sec. goal-box confinements.

## METHOD

*Subjects.*—The *S*s were 21 male albino rats of Wistar stock from the colony maintained by the psychological laboratory of Brown University. Their mean age was 84 days (range, 56–98) at the start of training. None had been used in prior experimentation.

*Apparatus.*—The apparatus was quite similar to one used by Sheffield (**4**). It consisted of a start box (9 × 6 in.), a straight alley (48 × 4 in.), and a goal box (12 × 8 in.). The inside height was 11 in. throughout. The start box was white; the remainder of the apparatus, black. Hinged glass lids covered the start box and goal box; the straight alley was covered by a hinged screen. The far end of the goal box opened into a food compartment containing a white food cup. The food compartment was small enough so that *S* could not enter it. Cord-

[1] This experiment was carried out by the junior author as an Honors project under the direction of the senior author who prepared this report.

[2] In an earlier paper (**6**), this type of extinction was labeled "frustrating" and distinguished from "nonfrustrating" extinction (e.g., extinction by omission of shock).

TABLE 1

SUMMARY OF PROCEDURE

| Day | Treatment | Hours Food Deprivation Prior to Treatment | Daily Food Consumption (gm. per Rat) |
|---|---|---|---|
| 1–4 | Taming: *S*s handled and fed in cages | 24 | 8 |
| 5–6 | Taming: *S*s handled and fed on table top | 24 | 7 |
| 7 | 10 training trials | 23–29* | 7 |
| 8 | 30 training trials | 22** | 6.5 |
| 9 | 30 extinction trials | 22** | — |

* Variations in food deprivation balanced in three extinction groups.

** After supplementary feeding which came about 1½ hr. after running.

operated sliding doors, one at the start box and the other 4 in. from the goal box, could be lowered to prevent retracing. A third sliding door could be lowered to block off the food compartment. A black curtain hung diagonally in the goal box.

A Springfield timer, which started when *E* raised the start-box door and stopped when *S*'s weight depressed the goal-box floor, gave a measure or response time which was recorded to the nearest .1 sec.

*Procedure.*—The schedule of taming, training, and extinction is shown in Table 1. Details of training were as follows: The start-box door was raised about 2 sec. after *S*'s feet hit the floor of the start box, and lowered as soon as *S*'s body cleared it. The goal-box door was lowered as soon as *S*'s weight depressed the goal-box floor. The *S* was left in the goal box for 10 sec. (longer if eating was slow), then placed in a between-trials cage for 20 sec. prior to its next trial.

A wet mash pellet made from ground Purina Lab Chow (average dry weight, .053 gm.) was used for reinforcement. During the first five training trials, the food compartment door remained open; during the remaining training trials, this door was raised only after *S* had entered the goal box and *E* had lowered the goal-box door. Two *S*s out of an original *N* of 24 were discarded for failure to eat in the goal box within 10 min. on some trial during training —a discard criterion set in advance. A third *S* died.

For extinction, each *S* was assigned at random to one of three conditions: (*a*) 10 sec. in goal box, 70 sec. in the between-trials cage (10-sec. group); (*b*) 30 sec. in goal box, 50 sec. in cage (30-sec. group); and (*c*) 60 sec. in goal box, 20 sec. in cage (60-sec. group). The *N* was seven in each group. During extinction, the food compartment door remained closed. If *S* failed to reach the goal box within 2 min., it was removed and retained in the between-trials cage for 80 sec. prior to its next trial.

## RESULTS

*Training.*—To check on the comparability of the three groups in performance during training, the median running times for successive 5-trial blocks were found for each *S*. The logs of the medians of these measures, for each group, are plotted in Fig. 1. It can be seen that the performances of the three groups were quite similar. No statistically significant differences obtained between any two groups when *S*s' median response times for each day and for blocks of five trials were analyzed with Wilcoxon's (7) nonparametric test for unpaired replicates.

*Extinction.*—It is apparent from Fig. 2 that differences in extinction conform to the hypothesis being tested. The 60-sec. group showed most extinction, the 10-sec. group least, the 30-sec. group fell in between. The 30-trial median response time difference between the 10-sec. and the 60-sec. groups is statistically significant, having a $p$ of .05–.02. This $p$ and those to follow include both tails, and hence are conservative.

One *S* in each extinction group failed to reach the goal box a sufficient number of times to make its over-all median determinate. In such cases, the *S* making less runs was ranked poorer. It is believed that this treatment does no serious violence to the assumptions underlying Wilcoxon's test.

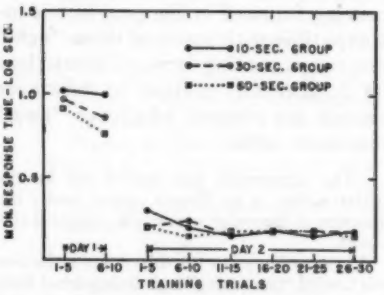


FIG. 1. Performance of the three extinction groups during training

Nonetheless, extinction was also analyzed in terms of a frequency measure of performance which has been used to treat comparable data (4, 5, 6). This measure was, for each *S*, the number of response times equal to or less than the over-all median response time of 17.0 sec. for all extinction trials for all *S*s. Again the 30-sec. group did not differ significantly from either of the other groups, but comparison of the 10- and 60-sec. groups in terms of this measure yielded $.02 > p > .01$.

The marked extinction "below zero" of the 60-sec. group is shown by the fact that the log median response times for each block of five extinction trials after the second block were significantly longer than those for the *first* training trial. The *p*'s are .05–.02 or better for all comparisons, according to Wilcoxon's paired replicates test. The 30-sec. group also showed differences which represent performance poorer than its unconditioned response time, but none of the differences is statistically significant. In contrast, the 10-sec. group barely returned to its unconditioned level of responding (Trials 11–15, Fig. 2), then improved.

Since all three extinction curves show reversals in direction during the middle part of extinction, statistical tests were carried out to determine whether these reversals can be attributed to chance. The over-all median response time for Trials 11–20 was determined separately for each group. Then for each *S* the number of responses faster than its group's median was found for Trials 11–15 and for Trials 16–20. Parametric *t* tests applied to the differences between these two frequency measures showed that the 60-sec. group and the 10-sec. group were responding more rapidly on Trials 16–20 than on Trials 11–15 (*p*'s of .02–.01). The *p* for the 30-sec. group was .64. The further comparison of the 30-sec. group's performance on Trials 6–10 with its performance on Trials 16–20 yielded a *p* of .10.

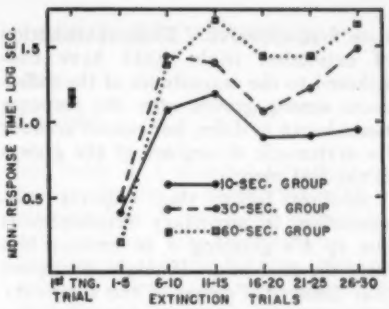


FIG. 2. Effect of goal-box confinement on performance during extinction

## DISCUSSION

The findings were as expected: the 60-sec. goal-box confinement group showed least resistance to extinction, the 10-sec. group most, and the 30-sec. group was in between. However, if these data are to be regarded as evidence for the interference theory of extinction described earlier, we must first show that our procedure adequately controlled distribution of extinction trials. This is especially necessary because Sheffield (5) and Stanley (6) have found that resistance to extinction as measured by running time in either a straight alley or a T maze was greater with massed than with spaced extinction trials.[3] It should be noted, therefore, that our extinction procedure assured that the time intervals between successive presentations of the conditioned stimulus—placement of *S* in the start box and opening of the start-box door—would be the same for all *S*s, unless systematic differences in response

[3] Both Sheffield and Stanley suggested that a "frustration-produced drive" or "excitement" factor due to the presence of stimuli associated with eating might account for their finding faster running during massed than with spaced extinction trials. More direct experimental verification of this frustration factor might throw light on the source of the variety and seeming vigor of responses that are observed during extinction by omission of food when hunger drive is present. It would also complicate, but not necessarily refute, an interference interpretation of the findings in the present study.

time first appeared. Thus, distribution of extinction trials could have contributed to the magnitudes of the differences among groups *after* the response times began to differ, but cannot account for systematic divergence of the groups in the first place.

Another factor that requires consideration is secondary reinforcement due to *E*'s grasping *S* to remove him from the goal box. It might be argued that differential delays in this secondary reinforcement could account for our findings since such delays were necessarily correlated with duration of goal-box confinement. This interpretation, however, is not an alternative interpretation, but one, we believe, that can be subsumed under an interference theory. Short goal-box confinements should not only prevent the occurrence of many competing responses, but also should render it more likely that *S* will be removed while making essentially the same response as on some previous extinction trial. If on a number of successive extinction trials the same response is followed by *E*'s removing *S*, then this response should become conditioned specifically to goal-box cues and disattached from start-box and stem cues. In short, recurrences of this response in the goal box should provide the basis for a new reinforcing property of goal-box cues. Although the compact running schedule precluded reliable recording of *S*s' behavior just prior to removal, *E* was impressed with recurrences of leaping in some *S*s. The recurrence of leaping was especially striking in several 10-sec. *S*s during the latter half of extinction. These *S*s ran into the goal box, waited until the goal-box lid was lifted, then leaped to the top edge of the goal box.

Whether the foregoing hypothesis can explain the improvement in performance after an initial decrement found in all three extinction groups, and clearly statistically significant in two of the groups, could be determined by an experiment in which *S*s were removed from the goal box only after the occurrence of some predetermined response. It might well be that after performance during extinction falls markedly below the unconditioned level of responding, as shown by the 60-sec. group, all that is required to produce some improvement in performance is that removals from goal box occur while *S* is locomoting and not after periods of quiescence in *S*.

Finally, it may be mentioned that the present findings have methodological significance, regardless of their theoretical import. If an investigator wishes to eliminate a conditioned running sequence to a point well below its unconditioned level of responding, it would be most economical to avoid very short durations of goal-box confinement. This information should be of interest to many investigators, considering the present wide usage of the speed of a rat's running during extinction in alleys and on runways for inferences concerning behavioral and physiological variables.

## Summary

From the observation that responses incompatible with forward locomotion typically occur in an "empty" goal box during extinction of a running response by omission of food, it was argued from contiguity theory that a long goal-box confinement should produce faster extinction than a short one. The hypothesis was tested and confirmed by training rats to run a straight alley, then carrying out extinction with 10-, 30-, and 60-sec. goal-box confinements. Theoretical and methodological implications of the findings were discussed.

## References

1. Estes, W. K. Toward a statistical theory of learning. *Psychol. Rev.*, 1950, **57**, 94–107.
2. Guthrie, E. R. *The psychology of learning* (Rev. Ed.). New York: Harper, 1952.
3. Sheffield, F. D. The contiguity principle in learning theory. *Psychol. Rev.*, 1951, **58**, 362–367.
4. Sheffield, V. F. Extinction as a function of partial reinforcement and distribution of practice. *J. exp. Psychol.*, 1949, **39**, 511–526.
5. Sheffield, V. F. Resistance to extinction as a function of the distribution of extinction trials. *J. exp. Psychol.*, 1950, **40**, 305–313.
6. Stanley, W. C. Extinction as a function of the spacing of extinction trials. *J. exp. Psychol.*, 1952, **43**, 249–260.
7. Wilcoxon, F. *Some rapid approximate statistical procedures*. Stamford, Conn.: American Cyanamid Co., 1949.

(Received February 9, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# A TEST OF ONE ASPECT OF CONTIGUITY THEORY[1]

A. D. DAVIS

*University of Washington*

Guthrie's theory of learning holds that temporal contiguity of stimulus and response is a necessary and sufficient condition for learning. The apparent effectiveness of reward as a determiner of learning is explained as being due to "removal from the situation." That is, response to the reward precludes further response to stimuli that have already entered into association with the rewarded responses, and thus preserves the "correct" associations in the same way that physical removal from the situation would. A direct test of this proposition involves removing the animal from the situation and observing whether or not the last response made before removal is learned. Removal may be physical (3) or it may involve removal of a class of stimuli (4). Results tend to be ambiguous because of the difficulty of saying, on the one hand, that removal was *immediate*, or, on the other hand, that removal had no primary or secondary rewarding characteristics. There are, however, implications inherent in the theory of learning by contiguity which lead to deductions that afford an indirect test, and such a test is presented here.

In applying contiguity theory to the learning of a simple T maze, one would expect that *S* would learn the incorrect turn after once having made that choice. As *S*s always tend to repeat the responses last made in a stimulus situation, contiguity theory is confronted with explaining how *S* learns the correct turn and "unlearns" the incorrect one. The problem is that of showing how outcomes of choices affect behavior at the spatially and temporally separated choice point where learning is displayed. In terms of contiguity theory this can only occur through the formation of associations involving stimuli which are operative both in the goal box and at the choice point. One way in which this may occur has been suggested by Guthrie and involves proprioceptive chaining. The relevant stimuli may be exteroceptive, interoceptive, or proprioceptive; this need not be specified for the present account, but for convenience of exposition we may assume that the stimulus consists of a light over the goal box which is visible at the choice point. According to the principle of contiguity, the light must become a cue for responses made in its presence and therefore must become associated with the approach responses made by *S* in making its first rewarded choice. By responding to the reward *S* is immediately removed from the stimuli in the maze and the approach responses are "preserved" intact as the last responses to that situation. On subsequent trials these responses are elicited by the light at the choice point and *S* is always led to the correct box.

In this manner the theory adequately explains how learning by strict contiguity of stimulus and response may be extended over spatial and temporal voids in fixing the correct response. Why then is the incorrect turn not similarly repeated after once having been made? In order to answer this question contiguity theory must necessarily hold that the last responses made after an incorrect choice are not conducive to repeating the choice. In contrast to the last responses to the lighted, rewarding choice, the incorrect choice results

[1] This paper is based on a thesis submitted to the faculty of the department of psychology of the University of Washington in partial fulfillment of the requirements for the M.S. degree. The author wishes to express his gratitude to Professor Moncrieff H. Smith for suggesting the problem and guiding the investigation until its conclusion.

in last responses that are random and variegated in direction. As these responses occur in the presence of darkness, they will be elicited on subsequent trials at the choice point if *S* orients toward the unlighted box. However, because of the indeterminate nature of these responses, they will not necessarily lead *S* back to the unrewarded choice. For example, the last responses to the unrewarded box are often those of escape, and when elicited on subsequent trials lead *S* to the alternative or correct box. These responses take precedence over the original associations of approach because of their later attachment to the stimulus of darkness (2, p. 96). In this way the incorrect turn is "unlearned" by the interference of these new, incompatible associations, and *S* learns the maze.

Now it becomes evident that such an explanation based on last responses may be investigated in a situation in which last responses are similar after both rewarded and unrewarded choices. If last responses of a variegated nature result in "unlearning" of unrewarded choices on subsequent trials, they should similarly affect rewarded ones. If *S* were retained in the goal box after reward, further interfering responses would be made and associated with stimuli (e.g., the light) which had previously been associated with approaching the reward. These last formed associations would now tend to be dominant and should interfere with the "correct" approach movements which are elicited at the choice point, slowing the learning of the correct turn. It is the purpose of this investigation to test this contiguity explanation by comparing learning with, and without, this element of delay after reward.

## Method

*Subjects.*—Forty-two male albino rats, 90 to 180 days old and experimentally naive, were used. They were divided into two groups of 21 rats each, as described below.

*Apparatus.*—The apparatus consisted of a conventional T maze with 30-in. arms, 6 in. in depth and 4 in. wide. The goal boxes were slightly modified by equipping them with drop bottoms. This modification was intended to standardize removal of all *S*s and insure removal at a given time. The only illumination in the room was a 7-w. bulb over the rewarded goal box. This was controlled by a double-throw, double-pole toggle switch mounted on the shield, making it possible to alternate the light dependent on the box to be rewarded.

*Procedure.*—All *S*s were randomly assigned to a control or experimental condition and to the box in which they were to be rewarded. The order in which *S*s received their one trial per day was held constant throughout the experiment in order to stabilize the 23-hr. deprivation from day to day. The reward consisted of a ½-gm. pellet of their regular laboratory chow diet. The control *S*s were removed from the goal box immediately after consuming the pellet, whereas *S*s in the experimental group were retained for 60 sec. after consuming the reward. This delay was determined by a pilot group and selected as one in which all *S*s had finished grooming and had engaged in further (interfering) responses. In order to equalize time spent in both boxes, the control *S*s were retained in the unrewarded box for 45 sec. in the event of a wrong turn and the experimental *S*s for 105 sec. The 45-sec. delay was mean consumption time for the ½-gm. pellet as determined by pilot study. All *S*s were run until the slowest had attained five successive correct choices.

## Results

In terms of the contiguity interpretation of interference from the delay, there is an antecedent expectation that the delayed *S*s' performance should differ from that of the nondelay group in the following respects: (*a*) The delayed group should show fewer successful (rewarded) choices; (*b*) they should require more trials to reach a level of proficiency; (*c*) they should display less "stereotypy" (tendency to repeat the choice of the previous trial); and (*d*) they should require more time to attach relevant cues to relevant responses, significantly increasing the running time of the maze.

The first three assumptions were tested by *t* tests of significance and yielded values of .53, 1.00, and .85, respectively. A *t* of 2.02 was needed for the .05 level of significance. The

mean latencies for the two groups were compared by graph and $t$ test ($t = .31$) and found to show no significant difference. As an additional precaution against overlooking differential group rates of learning, the number of animals per group making correct turns on successive trials is compared in Fig. 1. It is worthy of note that the levels of proficiency are comparable at any given trial as well as at termination. This further exemplifies the lack of difference between the groups also indicated by the statistical measure. However, it must be noted that these different approaches to the data are not independent measures.

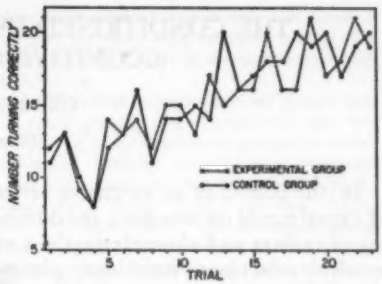


FIG. 1. Group performance on successive trials

## DISCUSSION

From the foregoing analysis it is apparent that the results are not in accordance with expectations from contiguity theory. There are at least two possible explanations from the contiguity position: (*a*) If the reward was insufficient to terminate the sequence of behavior which led to it, there would be no ensuing interference, merely a continuation of the sequence. In this instance, this explanation could have accounted for the equality of the groups only if there had been no learning. (*b*) Reward serves to terminate sequences and/or change the state of the organism sufficiently that subsequent acts are allied with a different stimulus pattern (1, p. 53). If this were the case in this investigation, the interference would have failed to develop as a result of delay after reward since the subsequent responses would have been allied with a stimulus state other than that associated with approaching the choice point. The random behavior in the empty goal box in this instance would get associated with the same stimuli which are operative at the choice point, whereas random behavior following reward would be associated with a different set of stimuli. The data presented here cannot answer this argument, but the results of other investigations indicate that it cannot be applied. For example, we can be fairly certain that the rats would still have learned had they been run under conditions of massed practice, rather than one trial per day. Furthermore, in the Skinner box an animal consumes a comparable reward and immediately returns to the lever for another. Both considerations suggest that within 1 min. after consumption of reward, the rat is returned to approximately the same stimulus state that had previously evoked the correct response.

## SUMMARY

It was shown that contiguity theory leads to the prediction that delay in the goal box after reward should retard learning. This deduction was tested by comparing performance in a T maze of a group removed from the goal box immediately after consuming the ½-gm. reward with a group retained in the box for one additional minute after reward. With 21 rats per group, there was no evidence of the difference predicted by contiguity theory.

## REFERENCES

1. GUTHRIE, E. R. On personality in terms of associative learning. In J. McV. Hunt (Ed.), *Personality and the behavior disorders*. Vol. I. New York: Ronald, 1944. Pp. 49–68.
2. GUTHRIE, E. R. *The psychology of learning*. New York: Harper, 1952.
3. SEWARD, J. P. An experimental study of Guthrie's theory of reinforcement. *J. exp. Psychol.*, 1942, **30**, 247–257.
4. ZEAMAN, D., & RADNER, L. A test of the mechanisms of learning proposed by Hull and Guthrie. *J. exp. Psychol.*, 1953, **45**, 239–245.

(Received March 1, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# THE CONDITIONED EVOCATION OF ATTITUDES (COGNITIVE CONDITIONING?)

GREGORY RAZRAN

*Queens College*

In the course of an extensive series of experiments on luncheon conditioning of ratings and characterizations of musical selections, paintings, photographs, and verbal statements (**9, 10, 11, 12**),[1] it was noted that, while the ratings and the characterizations could be significantly modified by the luncheon when *S*s were unaware of attempts to condition them, the conditioned stimuli by themselves, nonetheless, failed to evoke any special food or eating attitudes. The food or eating attitudes were inferred from, and measured by, three separate methods during the presentation of the conditioned material: (*a*) the frequency of food-related words and phrases in *S*s' free verbalizations about the conditioned material; (*b*) the frequency of food-related rhymes in *S*s' responses to a standardized list of 16 words that rhymed with both food-related and food-unrelated words; and (*c*) the speed with which *S*s unscrambled letter-scrambled food-related words. Each of the three methods was by itself quite effective in reflecting food attitudes between courses during lunch and immediately after lunch. Yet, none of them disclosed such attitudes when the conditioned stimuli alone were presented to *S*s a few hours after lunch, despite the fact that at the same time the same stimuli elicited reliable conditioned salivation. Apparently, what was conditioned in the series of experiments, besides salivation, was more a general affectivity than a specific attitude. Or, to put it differently, the conditioning was not redintegrative: it involved the conditioned transfer of some but not of all aspects of the unconditioned situation, and what was obtained was a conditioned *modification* rather than a conditioned *evocation* of attitudes.

However, in tabulating results, it was discovered that two *S*s failed, for some reason, to rate and characterize their to-be-conditioned musical selections, and these two did manifest considerable evidence of food attitudes during the postconditioned activations of the selections. Consequently, a hypothesis presented itself that *S*s' preoccupation with judging the experimental material was in some way responsible for the conditioning of general affectivities and not of specific attitudes; to test the hypothesis it seemed desirable to set up an auxiliary experiment in which no such judgments were to be made. Moreover, it was thought worth while to delay *S*s' luncheons until 2 o'clock in the afternoon so as to increase the intensity of the food attitudes; and in general, it appeared that the auxiliary experiment would further differ from the main experiment in at least four ways: (*a*) The durations of the experimental sessions would be reduced to one-third, since no comparisons of preluncheon and postluncheon judgments would be needed. (*b*) Not as many to-be-conditioned experimental

[1] The cited references are preliminary reports before APA and EPA meetings. The entire series is now being written up in book form as *Conditioned Attitudes: Likes, Dislikes, Judgments, and Beliefs*. The studies of the series were performed under the sponsorship of Professors Woodworth and Poffenberger, and were supported by the Council on Research in the Social Sciences of Columbia University.

items would be required, since, unlike the main experiment, there would be no question of preventing *Ss* from remembering specific judgments of specific items. (*c*) For the same reason—no question of remembering specific judgments—the number of luncheons per *S* (three in the main experiment) could be increased to allow a presumable increase in the strength of the conditioning. (*d*) Controlling *Ss*' attitudes toward the experiment, a perennial stumbling block in human conditioning, would be much less of a problem, since until the very end *Ss* would not be called on to react in any way. In short, the projected experiment was to be procedurally the simplest in *E*'s repertoire.

## Procedure

Twenty *Ss* (all college students) and 22 items of to-be-conditioned material were used (420 *Ss* and 132 items were employed in the main experiment). The 22 to-be-conditioned items consisted of (*a*) five short, but musically complete, excerpts of unfamiliar music; (*b*) five relatively unfamiliar paintings projected from color slides on a screen; (*c*) six photographs of college girls selected from various college yearbooks and projected on a screen to life size; and (*d*) two contrasting literary quotations, two controversial sociopolitical slogans, and two statements of fact—one true and one false. The musical selections were taken from Rameau's "Castor and Pollux" (orchestra), Chopin's "Mazurka No. 4, Opus 33" (piano), Debussy's "La plus que lente—Waltz" (violin), Benatzky's "Um das bisschen Liebe" (German jazz quartet), and "Udan Mas," a Javanese song (vocal). The paintings were: Gainsborough's "English Landscape," Koekkoek's "A Street in Holland," Degas' "After the Bath," Leger's cubist "The City," and a nonobjective Miro "Painting, 1927." The two literary quotations were "Be warm but pure; be amorous but chaste" and "Virtue flies when love once blows the sail," while the slogans were "Geographical pioneering is finished but social pioneering is just beginning" and "No government can spend itself prosperous." The presumably true statement was "Babies palm objects before they pick them up by finger-thumb opposition," and the presumably false statement was "Maternal behavior in the human female is by and large mostly inherited." (All verbal material was projected on a screen.) Finally, three of the photographs were of blonde girls and three of brunettes, and their mean rating for beauty on a 5-point scale by 350 *Ss* in the main experiment was $3.9 \pm .4$.

The 20 *Ss* in the experiment were divided into four equal groups, two experimental and two control. In both experimental groups the to-be-conditioned stimulus material was presented during six consecutive luncheons: Sat., Tues., Thurs., Sat., Tues., Thurs.; and Sat., Mon., Wed., Fri., Mon., Wed. The two experimental groups differed from each other in that one experimented with the music and the verbal statements and had lunch between 2:00 and 2:30, and the other experimented with the paintings and the photographs and had lunch between 2:30 and 3:00. The luncheons were in all cases paid for by *E*: two sandwiches, two desserts, and a beverage per *S*.[2] No food was, of course, given to the two control groups, who otherwise matched fully the experimental groups in pattern of stimulus presentation and in days and hours of experimentation (the control groups were not hunger groups, though; they had lunch at 12:00). The two varieties of stimulus material—music-statements or paintings-photographs—were each presented in both experimental and control groups twice in the same session; the individual items of the same variety followed each other in the same sequence and the two varieties alternated as *abba*.

Again, as in other experiments by *E*, *Ss* were misinformed about the true purpose of the experiment. They were told that *E* "was interested in finding out how different stimuli get connected up in one's mind and how eating affects such connections." The fact that two quite different kinds of stimulus material were used in the same session lent a great deal of extra credence to the explanation, and in no case was there any evidence that any *S* did not believe it. Indeed, the set direction—or misdirection—was so complete that in the second experimental session, *E*, fearing that preoccupation with interstimulus connections might interfere with building stimulus-food linkages, instructed *Ss* that they "should make no special effort to look for connections but merely to pay attention to the stimuli while concentrating on their luncheons and trying to enjoy it naturally and maximally." And later reports clearly indicated that the luncheon was by all signs a

[2] The *Ss* had their choice of kind of sandwich, dessert, and beverage; they were asked to bring to the laboratory whatever kind they liked best. But while there were, as might be expected, some variations in the choices of different *Ss* and of the same *S* at different times, no correlation was ever obtained in any study of the series between any index of conditioning and any food pattern of luncheon. The effective stimulus was at all times *S*'s choice rather than what he had chosen.

very pleasurable and natural—and meaningful—activity. The fact that it was paid for and the laboratory setting, if anything, heightened intensities of positive hedonic tones and magnitudes of motor-glandular involvements.

Finally, two weeks from the beginning of the experiment and a day after the last experimental session (Saturday) both the experimental and the control *Ss* were asked to come to the laboratory at 3 o'clock, and both were tested for the presence and magnitudes of food attitudes during the activation of their respective stimulus material. The tests mentioned at the beginning of the article were used: rhyme finding, word unscrambling, and free verbalization. The tests were, for convenience, divided among two separate sessions: the first session, from 3:00 to 3:40, was devoted to the rhyme finding and word unscrambling; and the second session, 3:50 to 4:30, was given to free verbalization. The stimulus material in the first session was presented twice so that each variety of material could be tested by each of the two methods. The stimulus material in the second session was presented only once thus affording about 3.5 min. of verbalization per individual item. The free-verbalization tests were merely analyses of *Ss'* written statements in response to the question: "What does the musical selection (or painting or photograph or statement) remind you of and what does it make you feel like doing?" The other two tests require special explanations.

Specifically, the rhyme-finding test consisted of 16 experimental words—beat, boast, born, cart, dice, fickle, haste, hunch, lake, lie, mood, mutter, real, shook, steam, and thread—which, as may be seen, rhyme with both food-related and food-unrelated words, and of 16 comparable control words that rhyme only with food-unrelated words. Its index of food attitudes was differences between conditioned and control *Ss* with respect to percentages of food-related rhymes in response to the experimental words during the presentation of the conditioned material. The word-unscrambling test contained, in its turn, eight scrambled food-related words—bhurrmaeg, digdpun, ledmat, malons, nooni, soottape, teelom, and thoocecal—and eight scrambled control food-unrelated words. Its index of food attitudes was differences between conditioned and control *Ss* in speed of unscrambling food-related words. Both tests were divided into two equal parts, one part for each of the two varieties of stimulus material, since there obviously would be no value in administering a specific word of either test more than once to the same *S*. Each part of the rhyme-finding test was, as might be expected, completed ahead of the time taken to present the conditioned material—5 or 6 items, 10 min. Each part of the word-unscrambling test was, as a rule, left unfinished. Again, it should be mentioned that in both tests, experimental and control words were randomly presented to prevent the formation of "spontaneous" food attitudes that might well have masked, if only partially, the conditioning results.

## Results

The results are all presented in Table 1, and, as may be seen, are quite definitive. The six columns under the "indices of conditioning" point to unmistakable differences between the conditioned and the control *Ss*. Nine of the 12 group differences were significant at the 1% level, and the remaining three were all in favor of the conditioned groups. The differences in *Ss'* unscrambling of scrambled food-related words seem to have been particularly outstanding—and interesting. These differences were significant at the 1% level for all the four varieties of stimulus material, and the over-all means disclose almost a 100% increase in the conditioned *Ss'* speeds of unscrambling. However, the conditioned stimuli were also quite efficacious in helping the conditioned *Ss* to produce larger proportions of food-related verbal material in their rhyme-finding and free-verbalization tests. While only two of four separate group differences in one case and three of four in the other case were significant at the 1% level, the over-all differences between conditioned and control groups were significant at this level in both cases.

Thus, it would seem, if we are permitted to generalize fully from not too large a sample, that when typical "human" stimuli—sayings, music, paintings, and photographs (and there is no reason for less "human" stimuli to behave too differently) are repeatedly activated (and again there is good reason to believe that repetition per se is but a secondary factor) while *Ss* are consuming an enjoyable meal, the stimuli thereby acquire a definite tendency to facilitate the evocation

TABLE 1

CONDITIONED EVOCATION OF FOOD ATTITUDES IN TEN ADULT *Ss*

Evocation of food-related verbal responses and facilitation of solutions of food-related verbal problems by stimuli that had been combined with six luncheons

| Conditioned Stimuli | Indices of Conditioning | | | | | |
|---|---|---|---|---|---|---|
| | Mean % of Food-Related Material in Free Verbalizations | | Mean % of Food-Related Rhymes | | Mean Speed (Min.) of Unscrambling Food-Related Words | |
| | Control *Ss* | Exp. *Ss* | Control *Ss* | Exp. *Ss* | Control *Ss* | Exp. *Ss* |
| 5 Musical selections | 4 | 12* | 34 | 43 | 2.9 | 1.8* |
| 5 Paintings | 3 | 9* | 33 | 48 | 3.0 | 1.4* |
| 6 Photographs | 3 | 7 | 28 | 46* | 2.6 | 1.4* |
| 6 Verbal statements | 2 | 10* | 22 | 41* | 2.7 | 1.3* |
| Over-all means | 3.0 | 9.4* | 29.3 | 44.8* | 2.8 | 1.5* |

* Difference between experimental and control groups significant at the 1% level. Percentages were used as scores in the first four columns, and *t* values were used to determine *p*'s in all columns.

of (or to evoke) food-pertaining verbal reactions. And inasmuch as such facilitations of specific reactions, which are well manifest in the three cardinal areas of verbal dynamics—free associations, controlled associations (rhyming), and problem solving—are presumably the ontological cores and the operational resultants of attitudes, we may well be justified in stating that conditioned attitudes were hereby evoked.

## DISCUSSION

To *E*'s knowledge, no experiment on conditioned evocation of attitudes has ever been reported. Conceptually, the experiments close to it in form seem at first to be those called "sensory conditioning" by Cason (6); they presumably also involve the evocations of central and intraorganismic events rather than of peripheral effector reactions. However, in sensory conditioning, the central event, the sensation, while behaviorally implicit, is of course experimentally very explicit during both the original training and subsequent testing. But in the present experiment, explicit hunger sensations were, if at all, only a secondary factor in the total psychophysiological matrix of the food attitude—in its function as a facilitating-mediating and directing-controlling variable. Indeed, when at the end of the experiment, *Ss* were specifically asked "what *extra* sensations the conditioned stimulus material aroused in them," the positive answers (not too many) were mostly descriptions of mouth waterings—induced by the conditioned salivations—and not of hunger sensations. By all signs the behaviorally re-evoked food attitudes were attendant by but little reinstatement of conscious processes, and in this respect their conditioning was really much like that of unconscious autonomic activities—as in the work of Bykov and his coworkers (3, 4, 5). But, again, autonomic conditioning has so far dealt primarily with the conditioned re-evocation of specific reactions—mostly peripheral—and with material that has never been, or could not be, in consciousness, rather than with what has been there at one time but is no longer reported as such. In short, what was conditioned here—what the unconditioned reaction was—was neither a sensation nor a motor nor a glandular nor a verbal reaction but something like a *Determinierende Tendenz*, which Ach in 1935 (2), as in 1905 (1), regarded as basically unconscious, and the core of which, even Titchener thought, could be "carried in purely physiological terms" (15, p. 369). Or, in other words, what was obtained here was *cognitive conditioning*.

It will also be remembered that it was stated at the beginning of the report that

this cognitive conditioning, or conditioning of food attitudes, was not realized when *Ss* were engaged during the training trials in what might be called another dominant activity, namely, rating and characterizing the to-be-conditioned stimuli. Lashley's view that "In any trial of a training series, only those components of the stimulating situation that are dominant in the organization [of any nervous activity or reaction] are associated" (7, p. 242) was thus corroborated. And it should be added that findings accordant with this view are clearly evident in all of the writer's conditioning experiments: conditioning priority of simple tone ratios, configured light patterns, and semantic aspects of words and sentences (8, 13, 14). However, in almost all cases, "dominance determination" finally resided more in *Ss*' attitudes than in the mere "compellingness" of the stimuli. When the to-be-conditioned stimuli, for instance, were flashes of miniature red and green lights which at the same time served as cues in solving a bolt maze, the course of the conditioning was much more a function of "rightness" or "wrongness" of the lights than of their spatial arrangements. Again, in the writer's experiments, "what was conditioned" was not always just a matter of "all or none" but also of "more and less." True, no "helter-skelter" of excitations but only organized aspects of situations were parts of the conditioning equation. But more than one aspect could at times be conditioned in varying degrees, and certainly more than one response could be so affected. In the present experiment both *Ss*' cognitive food attitudes and their reflexive salivary responses were conditioningly modified.

## Summary

1. When musical selections, paintings, photographs of college girls, and quotations and slogans were presented to ten adult human *Ss* during six consecutive free luncheons, the presented material came to evoke thereby conditioned food attitudes (*Ss* were misinformed about the true nature of the experiment).

2. Food attitudes were determined by frequencies of food-related free verbalizations, frequencies of food-related rhyme finding, and speed of unscrambling food-related letter-scrambled words. In all three indices the differences between experimental and control *Ss* were statistically significant.

3. Reported hunger sensations during the postconditioned no-food presentations of the experimental material were few and statistically insignificant. It is thought therefore that what was conditioned in the study was more an unconscious cognition in the sense of Ach's *Determinierende Tendenzen* (which Ach and even Titchener regarded as basically unconscious) than a conscious sensory or phenomenal experience.

## References

1. Ach, N. *Ueber die Willenstätigkeit und das Denken.* Gottingen: Vanderhoeck & Ruprecht, 1905.
2. Ach, N. *Analyse des Willens.* Berlin and Wien: Ruban & Schwartzenberg, 1935.
3. Bykov, K. M. The functional connections of the cerebral cortex with the viscera. *Acta Med. URRS.*, 1938, **1**, 178–203.
4. Bykov, K. M. *The cerebral cortex and the internal organs.* Moscow: Medgiz, 1947.
5. Bykov, K. M. The doctrine of reflexes and conditioned reflexes. *Fiziol. Zh. SSSR.*, 1950, **36**, 394–403.
6. Cason, H. Sensory conditioning. *J. exp. Psychol.*, 1936, **19**, 572–591.
7. Lashley, K. S., & Wade, M. The Pavlovian theory of generalization. *Psychol. Rev.*, 1946, **53**, 72–87.
8. Razran, G. Attitudinal control of human conditioning. *J. Psychol.*, 1936, **2**, 327–337.
9. Razran, G. Music, art, and the conditioned response. *Psychol. Bull.*, 1938, **35**, 532. (Abstract)
10. Razran, G. Conditioning away social bias by the luncheon technique. *Psychol. Bull.*, 1938, **35**, 693. (Abstract)
11. Razran, G. Conditioned response changes in rating and appraising sociopolitical slogans. *Psychol. Bull.*, 1940, **37**, 481. (Abstract)
12. Razran, G. Determinants of the consolidation or *Praegnanz* of conditioned preferences. *Psychol. Bull.*, 1940, **37**, 564. (Abstract)
13. Razran, G. Stimulus generalization of conditioned responses. *Psychol. Bull.*, 1949, **46**, 337–365.
14. Razran, G. Experimental semantics. *Trans. N. Y. Acad. Sci.*, 1952, **14**, 171–177.
15. Titchener, E. B. *Lectures on the experimental psychology of the thought-processes.* New York: Macmillan, 1909.

(Received March 1, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# SET OR PERCEPTUAL DEFENSE?[1]

JAMES T. FREEMAN[2]

*Northwestern University*

Several theoretical papers (**3, 4, 9, 10**) and recent experimental evidence (**1, 5, 11**) have suggested alternative explanations of the data previously explained in terms of perceptual defense (**6, 7**). The phenomenon of increased tachistoscopic recognition thresholds to emotionally-toned stimuli (perceptual defense) may, it appears, be a function of variables of a simpler order than those required by a "regressive-type" defense mechanism.

It has been suggested (**3**), for instance, that the perceptual-defense phenomenon may represent a tendency to withhold the verbal report in a selective manner in a compromising experimental situation. The *S*'s predominant set may predispose him to try some alternative hypothesis, not the correct one, which necessarily results in an increased recognition threshold (**12**). In a situation in which the number of alternative hypotheses is reduced, e.g., by an appropriate set to expect emotionally-toned material, the perceptual-defense phenomenon should be virtually eliminated.

The purpose of the present research is to determine the influence of a set created by instructions upon the perceptual-defense phenomenon.

[1] This study is based upon a portion of a dissertation submitted to the Graduate School of Northwestern University in partial fulfillment of the requirements for the Ph.D. degree. The writer would like to express his appreciation to Dr. Carl P. Duncan, who directed the research, and to Drs. B. J. Underwood and Irwin A. Berg for helpful suggestions made in the course of the study.

[2] Now at Iowa State College.

## General Method

*Apparatus.*—The apparatus used was the Gerbrands Mirror Tachistoscope, which allows for controlled exposure of stimulus words by means of a timer calibrated in intervals of .01 sec. from .01 to 1.00 sec.

*Subjects.*—The *S*s were 88 male students from the introductory psychology classes at Northwestern University. Each *S* participated in two experimental sessions 24 hr. apart with each session lasting from 15 to 30 min. All *S*s were asked not to discuss the experiment with anyone, and an attempt was made to ascertain before the experiment started whether or not *S* had received any information about the experiment. All *S*s used were naive with respect to the type and purpose of the experiments.

*Stimulus materials.*—Two types of stimulus words were used, neutral words that were taken from the Thorndike-Lorge general word count (**13**) and that were equated with respect to frequency and length, and taboo words that were of a sexual or scatological nature. The taboo words were equated with the neutral words with respect to length. Frequency of usage, of course, could not be determined. The stimulus words were typed in capital letters upon adding-machine paper and were exposed through an aperture in the back of the apparatus.

*Procedure.*—The two sessions consisted of a practice session to acquaint *S* with the procedure and to serve as a basis for matching *S*s, and an experimental session in which *S* received the conditions appropriate to either the experimental or control group. Each experiment made use of an experimental group and a control group. The conditions for the two groups differed only with respect to the type of instructions given. The two groups were matched on the basis of the mean recognition thresholds on the practice day for five four-letter words (neutral) equated for frequency of usage.

On the practice day all *S*s were given the same instructions and the list of stimulus items consisted of neutral words only. On the experimental day the instructions differed for the experimental and control groups and the list consisted of both taboo and neutral words.

A modified method of limits (ascending order) was used in presenting the stimulus words. The points of origin were varied in order to minimize errors of expectation. The initial

exposures were well below threshold and the series was continued until *S* made two consecutive veridical reports. The threshold was defined as the duration of the final exposure. Prerecognition hypotheses were recorded for each exposure. All *S*s who had not reached a recognition threshold by 1.00 sec. were discarded since the tachistoscope was not calibrated beyond this range.

Half of each group in each experiment received the stimulus words in one order and the other half in exactly the reverse order. This reduced any systematic bias due to practice effect. The list was so arranged that two taboo words never occurred in succession and a taboo word was never first in the list. With this restriction imposed, the serial position of the taboo words was, in fact, the same for all *S*s.

Before the exposures were started in each session, *S*s were given time to get adapted to the illumination in the tachistoscope and to get comfortably situated before the apparatus. There was no limit to the time for experimental sessions nor to the time in which *S* had to respond.

## Experiment I

*Method.*—This experiment was concerned with the recognition thresholds to taboo (and neutral) words as a function of the type of instructions. A six-word list containing two taboo words was used. All words were four letters long. The four neutral words were "lair," "omen," "wiry," and "numb." The instructions to Control Group 1 were of a general type, i.e., that words would be presented and that they were to respond with what they thought they saw. The instructions to the experimental group were to the effect that they were to be presented a list of words, some but not all of which were taboo words, and that they were to respond as soon as possible to all words. They were given examples of what was meant by taboo words and were told to feel perfectly free to respond with whatever word they thought they saw. This last procedure was an attempt to create a "permissive atmosphere." An additional control group (Control Group 2) was run using the same words but with instructions that the list was to contain some words relative to types of fruits and that they were to respond to all words as soon as possible. This control was used to provide data relative to the effect of a set (although inappropriate) to look for something specific which would not be present in Control Group 1 and which might exert some lowering effect upon thresholds to all words. Each of the three groups contained 16 *S*s.

*Results.*—The data are reported in mean recognition thresholds in seconds. Since the groups were matched, all tests of significance are based upon correlated measurements.

The mean threshold for all words in the *practice* list was .116 for Control Group 1, .117 for Control Group 2, and .114 for the experimental group. None of the differences among these means is significant; the groups are well matched.

The data for the taboo words in the *experimental* list are presented in Table 1, and for all words in the list in Fig. 1. The differences between the control groups on Stimulus Word 1 (neutral), and Words 2 and 5 (taboo) were tested for significance. None of the differences approached significance, indicating that instructions to look for something specific did not reduce the thresholds compared to the general instructions. Therefore, the two control groups were combined into a single group to compare with the experimental group. Table 1 shows that the experimental group had a significantly lower threshold for both taboo words. Although the differences for the second taboo word (Number 5) are less than for the first (Number 2), both differences are

TABLE 1

Mean Thresholds of Taboo Words and Results of *t* Tests between Groups in Exp. I

| Group | *N* | Serial Position of Taboo Word | | | |
|---|---|---|---|---|---|
| | | 2 | | 5 | |
| | | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| Control 1 and 2 | 32 | .199 | .024 | .121 | .015 |
| Experimental | 16 | .068 | .006 | .063 | .006 |
| *t* | | 5.03** | | 3.41** | |

** Significant at the .01 level.

highly significant. Since $F$ tests indicated heterogeneity of variance between the groups for both taboo words, the approximation $t$ test described by Edwards (2) was used. It may be noted that the thresholds for the experimental group are, in general, lower than the thresholds for the control group. This would seem to indicate that specific instructions (or sets) when confirmed early by the stimulus information serve to reduce variability (i.e., significant differences in variance), for these *S*s appear to be functioning somewhere near the lower asymptote of perceptual recognition and, indeed, show a certain "sensitization" to the taboo words.

In line with the above analysis, the thresholds for both taboo words were combined, and a comparison was made between the experimental and control group. The resulting $t$ of 5.57 was highly significant. It may be observed in Fig. 1 that the thresholds to the taboo words were lower than the thresholds to the neutral words for the experimental group (mean difference = .030). This difference could be due, in part, to the fact that the experimental *S*s had a more restricted range of hypotheses from which to work. There are a relatively small number of four-letter taboo words and, in the absence of marked preferences, any given hypothesis had a greater probability of being confirmed.

There appears to be a trend between the first and second taboo word in the control group toward a lower threshold. A test was made of the significance of this difference for the combined control groups; a $t$ of 3.08 was obtained. A further check was made to determine the extent to which general practice effects may account for such a trend. By using the practice-day curve as a base line, it was found that approximately 15% of the change could be attributed to general practice effects. A change of such magnitude did not appreciably alter the significance levels of the differences, however. It appears that there is a specific habituation effect operating between the first and second taboo word in the control group.

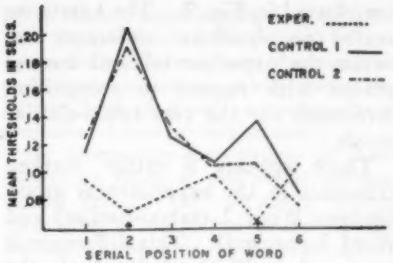


FIG. 1. Mean recognition thresholds as a function of stimulus word in Exp. I. Arrows indicate taboo words.

The results of Exp. I show no evidence for perceptual defense against taboo words when *S*s have been set by instructions to look for and report such words.

## EXPERIMENT II

*Method.*—An experimental group and one control group ($N$ = 20 each) were used. The instructions to each were the same as those used for the experimental group and Control Group 1 in Exp. I. The list of words was the same as in Exp. I except that the first letter of each of the two taboo words was changed, making them English words. These taboo-similar words were "hiss" and "muck." They have the same frequency of occurrence (five per million) as the four neutral words. Since the instructions to the experimental group are that the list contains some taboo words, the purpose of Exp. II is to test for the effect of such instructions on words which are taboo-similar.

*Results.*—The mean threshold for all words in the practice day list was .130 for the control group and .126 for the experimental group; the difference does not approach significance.

The data for the experimental list

are plotted in Fig. 2. The *t* tests revealed no significant difference between the experimental and control groups with respect to recognition thresholds for the two taboo-similar words.

There appears a rather marked difference in the experimental group between Word 2 (taboo-similar) and Word 3 (neutral). This difference is presented in Table 2 for both the control and experimental groups. The difference in the case of the latter is significant at the .02 level, although heterogeneity of variance was present and the approximation *t* test was used.

A further check was made of *Ss'* protocols to determine whether or not there was any relationship between the increase in threshold to the item following the taboo-similar word and the nature of the prerecognition hypotheses to the taboo-similar word itself. The total number of taboo prerecognition hypotheses, i.e., instances where *S* actually called out a taboo word preceding correct recognition, was tabulated for both groups. Also a tabulation was made of the total number of taboo prerecognition hypotheses which occurred to the first taboo-similar word only. These data appear in Table 3. Chi-square tests were performed, and the differences between the number of taboo prerecognition hypotheses for the experimental and control groups appear significant in both cases. Furthermore, it appears that the recognition thresholds for the experimental group remain uniformly higher than for the control group throughout the remainder of the list with the exception of the very last word.

TABLE 2

MEAN THRESHOLDS OF WORD 2 (TABOO-SIMILAR) AND WORD 3 (NEUTRAL) AND RESULTS OF *t* TESTS BETWEEN GROUPS IN EXP. II

| Stimulus Word | Group | | | |
|---|---|---|---|---|
| | Control | | Experimental | |
| | Mean | $\sigma_M$ | Mean | $\sigma_M$ |
| 2 | .115 | .019 | .103 | .012 |
| 3 | .130 | .014 | .166 | .022 |
| *t* | 1.07 | | 2.62* | |

* Significant at .02 level.

In this experiment false instructions that taboo words would be presented did not affect the thresholds of taboo-similar words but did increase the number of taboo prerecognition hypotheses and raise the thresholds of neutral words following the first-appearing taboo-similar word.

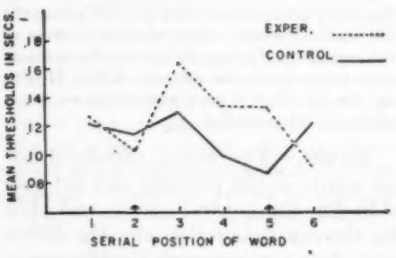


FIG. 2. Mean recognition thresholds as a function of stimulus word in Exp. II. Arrows indicate taboo-similar words.

## DISCUSSION

As a whole, the results indicate that factors related to set are important determinants of perceptual thresholds. In an experimental situation which has customarily produced evidence for perceptual defense, no evidence for this phenomenon was found when set was suitably manipulated by instructions.

In Exp. I, it was demonstrated that threshold differences to taboo words could be significantly reduced by introducing instructions which prepared *S* both to expect taboo words and to feel free to verbalize what he saw. Similar results have been obtained by Lacey, Lewinger, and Adamson (5), who varied foreknowledge relative to the presentation of taboo words. A "regression

effect" similar to that obtained between the first and second taboo words in Exp. I was also noted by these investigators. That is, upon successive presentations of taboo words the thresholds tend to become lower, thus revealing a type of habituation effect to these words. This effect was also noted by Bitterman and Kniffen (1) in an investigation of the relationship between manifest anxiety and perceptual defense. These writers concluded that perceptual defense appears to be related to differential readiness to report, determined by personality factors of the various *S*s. It was further observed in our Exp. I that the experimental group not only had lower thresholds to the taboo words than did the control groups but that they had lower thresholds to the neutral words as well. This would seem to indicate that the effect of set or a hypothesis that is confirmed early in the series is to lower thresholds in general as compared to no specific set (Control Group 1) or a set that is not confirmed (Control Group 2). However, the experimental group also had lower thresholds to taboo as contrasted to neutral words indicating a still more specific function of set.

Not only can the factors producing what has been characterized as perceptual defense be manipulated by instructions, but the results of Exp. II suggest that the "perception which is being defended against" can actually be induced by set and the use of taboo-similar material. The results of Exp. II are not consistent with the implications of perceptual defense, for it is unreasonable to say that stimuli were being defended against at a time when *S* was freely verbalizing taboo prerecognition hypotheses as to the nature of the stimuli. In addition, there was no evidence for increased thresholds to the taboo-similar words as might be expected if a mechanism such as perceptual defense were operative. However, a type of "defense" or more properly a "vigilance" does seem to develop to all words following the taboo-similar word which *S* has incorrectly perceived as a taboo word. (Lacey, Lewinger, and Adamson [5] also employed a similar procedure using taboo-similar words, but no data were reported for these words.) Once again, the pervasive effects of set may be noticed in an instance where words are indeed "misperceived" according to a particular dominant hypothesis.

An alternative formulation of these results, in place of perceptual defense, would explain the ordinarily higher thresholds of the taboo words (in the traditional perceptual-defense situation) as a function of the dominance of alternative sets which do *not* predispose *S* toward the perception of taboo material. Such an explanation is consistent with other experimental evidence and may readily be subsumed under an already existing proposition (8) which asserts, in effect, that the strength of a given hypothesis is a function of the number of alternative hypotheses available. For example, in the perceptual-defense situation a number of hypotheses may be available at the outset. Therefore the probability of any one being confirmed, e.g., a hypothesis for taboo words, is less than in a situation in which the number of possible hypotheses is reduced, e.g., by instructions to expect taboo words.

It seems quite unlikely that *S* initially has a strong hypothesis that he will be presented taboo material. In our Exp. I, in particular, it would seem that on the experimental day the control group *S*s would have a general set toward

TABLE 3

FREQUENCY OF TABOO PRERECOGNITION HYPOTHESES FOR THE FIRST-APPEARING TABOO-SIMILAR WORD, AND COMBINED FREQUENCY FOR BOTH TABOO-SIMILAR WORDS IN EXP. II

| Group | *N* | First Taboo-Similar Word | | Both Taboo-Similar Words | |
|---|---|---|---|---|---|
| | | Observed | Expected | Observed | Expected |
| Experimental | 20 | 10 | 6 | 19 | 11 |
| Control | 20 | 2 | 6 | 3 | 11 |
| $\chi^2$ | | 4.08* | | 8.95** | |

* Significant at the .02 level.
** Significant at the .01 level.

neutral material since they had had experience with this material on the practice day and had received no instructions to expect any different material on the experimental day. Thus, at the outset of the experimental session, they were likely to be set for neutral material, and therefore their thresholds to taboo words were higher relative to their thresholds for neutral words. On the other hand, the experimental group had a specific set from the instructions toward taboo words on the experimental day, and thus their thresholds to these words were lowered. In other words there were, in effect, two different sets operating with respect to the two groups: in the experimental group the set or hypothesis for taboo material was confirmed and lower thresholds to these words resulted, whereas in the control group, only after the dominant hypothesis or set (an expectancy of more or less neutral material) had first been invalidated did correct recognition of the taboo words occur. This necessarily resulted in increased thresholds.

## Summary

Two experiments were performed to determine the influence of set upon a phenomenon which has been described as perceptual defense. The experiments involved tachistoscopic presentation of taboo, taboo-similar, and neutral words. The variables investigated were: (*a*) the effect of instructions preparing *S* for taboo words, and (*b*) the effect of preparing *S* for taboo words but presenting taboo-similar words. The *S*s were male undergraduate students matched on the basis of a practice day performance. The measures used were recognition thresholds (time) and prerecognition hypotheses. The data support the following conclusions:

1. Conditions producing apparent perceptual defense may be manipulated by varying the instructions given to *S*. Where *S* is set to expect taboo material, perceptual defense does not occur.

2. Where *S* is set to expect taboo words, a sensitization to these words occurs. Where *S* is not originally set to expect taboo words, a habituation effect to successive taboo words occurs.

3. The *S*s may be induced, through set, to report taboo prerecognition hypotheses to words similar to taboo words with a subsequent "vigilance" effect to the following words.

It was suggested that the results provide a more direct explanation of the data which have previously been interpreted in terms of a mechanism of perceptual defense. Specifically, a perceptual principle of greater generality which would relate the ordinarily increased recognition thresholds of taboo material to the dominance of alternative hypotheses or sets was proposed.

## References

1. Bitterman, M. E., & Kniffen, C. W. Manifest anxiety and "perceptual defense." *J. abnorm. soc. Psychol.*, 1953, **48**, 248–252.
2. Edwards, A. L. *Experimental design in psychological research.* New York: Rinehart, 1950.
3. Howes, D. H., & Solomon, R. L. A note on McGinnies' "Emotionality and perceptual defense." *Psychol. Rev.*, 1950, **57**, 229–234.
4. Howie, D. Perceptual defense. *Psychol. Rev.*, 1952, **59**, 308–315.
5. Lacey, O. W., Lewinger, N., & Adamson, J. F. Foreknowledge as a factor affecting perceptual defense and alertness. *J. exp. Psychol.*, 1953, **45**, 169–174.
6. McGinnies, E. Emotionality and perceptual defense. *Psychol. Rev.*, 1949, **56**, 244–251.
7. McGinnies, E., & Sherman, H. Generalization of perceptual defense. *J. abnorm. soc. Psychol.*, 1952, **47**, 81–85.
8. Postman, L. Toward a general theory of cognition. In J. H. Rohrer & M. Sherif (Eds.), *Social psychology at the crossroads.* New York: Harper, 1951. Pp. 242–272.
9. Postman, L. On the problem of perceptual defense. *Psychol. Rev.*, 1953, **60**, 298–306.
10. Postman, L. Perception, motivation, and behavior. *J. Pers.*, 1953, **22**, 17–31.
11. Postman, L., Bronson, W. C., & Gropper, G. L. Is there a mechanism of perceptual defense? *J. abnorm. soc. Psychol.*, 1953, **48**, 215–224.
12. Postman, L., & Bruner, J. S. Multiplicity of set as a determinant of perceptual behavior. *J. exp. Psychol.*, 1949, **39**, 369–377.
13. Thorndike, E. L., & Lorge, I. *The teacher's word book of 30,000 words.* New York: Columbia Univ. Press, 1944.

(Received February 23, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# THE EFFECT OF STIMULUS COMPLEXITY ON DISCRIMINATIVE RESPONSES[1]

LEE W. GREGG[2]

*University of Wisconsin*

Control systems utilizing visual displays require rapid, error-free responses on the part of the operator for optimum use. The responses depend on the detection of differences in the stimulus objects of the display. The present investigation provides information relative to the nature of the changes in proficiency of performance of discriminative responses to visual stimuli that vary in location, size, or intensity.

Since visual displays usually incorporate more than one dimension of variation in them so that as much information as possible can be presented in a single unit, a related problem is the determination of the effects of added dimensions on discriminations in a first dimension. A second purpose of this study is to present a method of scaling stimuli in terms of a complexity criterion and show how proficiency of performance of the discrimination task varies as a function of this variable. The quantification of stimulus complexity is carried out independently of the particular dimensions of stimulus variations within the sensory mode, and consequently it may prove to be useful in comparing performances during acquisition over a wide range of stimulus conditions. However, the performance measures in a discrimination task do not depend on the nature of the stimuli alone. Response variables such as the type of control, kind or direction of movement, extent of movement, and the like determine the limits within which the proficiency measures can vary. Moreover, an interaction effect, or relationship extending over both the stimulus and response characteristics, exists as a separate variable that influences proficiency measures. This variable is referred to as stimulus-response compatibility, and the evidence for the existence of such a variable has been summarized by Fitts (1, p. 1306).

The quantification of stimulus complexity is derived from the following considerations. Stimuli may be described in terms of physical properties attributed to them. Such properties as duration and intensity are common to all stimuli and are continuously variable. The exact nature of any stimulus within a particular sensory mode depends upon the relative amounts or levels of such dimensional properties. In order to differentiate between any two stimuli, there must exist a difference in at least one dimension common to the class of stimuli. If the difference in a single dimension

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-23294 with the University of Wisconsin Laboratory of Experimental Psychology, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, publication, use, and disposal in whole or in part by or for the United States Government. The article is based on a thesis submitted to the faculty of the Graduate School of the University of Wisconsin in partial fulfillment of the requirements for the degree of Doctor of Philosophy. The author expresses his appreciation to Dr. W. J. Brogden under whose direction the research was carried out. He also wishes to thank the other members of the research committee, Drs. D. A. Grant and F. A. Mote, for their guidance and many helpful suggestions.

[2] Now at Carnegie Institute of Technology.

exists, and is such that the two stimuli to be compared are differentiated, and appropriately responded to every time the stimuli are presented, there exists the simplest condition of stimulus complexity. However, the two stimuli may vary within the two dimensions so that there is no correspondence between the levels of the two dimensions. Under these conditions, the equally discriminable differences in the two dimensions, over a long run, provide the bases for two independent discriminations with which the differential response may be associated. This represents a case where the correct association of the stimuli and the responses would be predicted to be slower in its establishment than in the case where only a single dimension is varied and only one basis for distinguishing the stimuli exists. Consequently, greater stimulus complexity is inferred from these hypothesized results. Since the increase in complexity is supposed to result from the addition of one more stimulus dimension, the stimulus complexity is said to be increased by one complexity unit.

A rational basis for such a scaling procedure and a formula for the derivation of the scale of stimulus complexity are possible if the stimuli are considered as sources of information. A recent theory of communication, developed by Shannon (7), of which information measure is an integral part has already found several applications in psychological research (2, 6). The concept of the amount of information is applied here only to the specification of the scale values of stimulus complexity.

In the discrimination-learning problem, two stimuli are presented either successively or simultaneously, and these stimuli require some differential response. If the stimuli differ only in terms of the difference in the level of a single dimension, the learned association is that of relating Stimulus $s_1$ with Response $r_1$ and $s_2$ with $r_2$, or that of relating $s_1$ with $r_2$ and $s_2$ with $r_1$. There are, then, only two possible states or choices. This particular situation has already been specified as the simplest stimulus complexity condition. The expression used by Shannon (7) to define the amount of information, $I$, for an operating cycle is:

$$I = \log_2 (\text{number of possible states of the source}).$$

The unit is the binary digit or bit. If this formula is applied to the stimulus complexity dimension, it is readily seen that adding dimensions, each with two levels, one at a time yields 2, 4, 8, and 16 "states" as we go from one to four dimensions. The $\log_2$ computation for these stimulus complexity conditions gives scale values of 1, 2, 3, and 4 bits, respectively.

Visual stimuli were used in the present investigation and varied to provide stimulus conditions corresponding to scale values of stimulus complexity of 1, 2, 3, and 4 bits. The discriminative responses were movements of a joy-stick control transversely to the right or to the left. Four dimensions of variation of the visual stimuli were chosen: the spatial location in the horizontal plane, the spatial location in the vertical plane, size, and intensity. For the 1-bit scale value of stimulus complexity, the stimuli vary in only one dimension; e.g., the separate stimuli are either left or right. For the 2-bit scale value of stimulus complexity, the stimuli vary in two dimensions, as right or left and up or down. The stimuli vary in three or four dimensions for the 3- and 4-bit scale values. To meet the requirements of the hypothesized scale, the

right-left, up-down, large-small, and bright-dim differences were such that they were always detected. Also, the occurrence of the levels of the dimensions when two or more were combined were random—each level of the separate dimensions occurring equally often with each level of every other dimension. As a parameter for the stimulus complexity function, four stimulus-response compatibility relations were established. The actual responses were the same, but the right or left movement was to correspond with the right-left, up-down, large-small, or bright-dim stimulus variation for the separate group treatments.

## Method

*Experimental design.*—Four groups of 24 *Ss* each were used. The *Ss* of Group I were instructed to move the joy stick to the right when the target appeared to the right of the screen, and to the left when the target appeared to the left. The *Ss* of Group II were instructed to move the joy stick to the right when the target was up, and to the left when the target was down. Those of Group III were instructed to move the joy stick to the right when the target was large, and to the left when the target was small. Finally, Group IV *Ss* were instructed to move the joy stick to the right when the target was bright, and to the left when the target was dim.

The experimental session consisted of an acquisition phase and a transfer phase, the latter devoted to the evaluation of the effect of stimulus complexity on the proficiency measures for the discriminative responses. The acquisition phase provided 48 trials under the stimulus condition corresponding to the simplest complexity value, i.e., the targets varied only within a single dimension. Rest periods of approximately 2 min. followed Trials 24 and 48. After the acquisition trials, four blocks of 24 trials each were given, with 2-min. rest periods between successive blocks. Each block of trials represented one of the four stimulus conditions corresponding to the four levels of stimulus complexity. The order in which these stimulus conditions were presented was determined by the rows of a $4 \times 4$ latin square. The same square was used for all groups. Within each group, six *Ss* were randomly assigned to each of the four sequences. Since there were six essentially distinct stimulus patterns for the stimulus condition corresponding to the 3-bit scale value of stimulus complexity, the 24 *Ss* within each group constituted but a single replication of the design.

The specific stimulus conditions were as follows. For Group I, Cond. 1 (scale value of stimulus complexity = 1 bit) consisted of two separate stimuli or targets that were different only in their spatial location in the horizontal or right-left plane. The spatial separation from center to center of these targets, circular light patches .25 in. in diameter, was 1 in. For the stimulus condition corresponding to the 2-bit value of stimulus complexity, one additional dimension was added to the right-left variation. In order to balance any specific effects, all three of the remaining dimensions were used, one at a time. Thus, for Group I, Cond. 2 actually consisted of three stimulus patterns: right or left targets that were also either up or down; right or left targets that were also either large or small; and right or left targets that were also either bright or dim. For Group I, Cond. 3 had six patterns for the combinations of the dimensions: (*a*) right or left stimuli that were also either up or down and large or small, but were right or left and up or down for Cond. 2; (*b*) right or left stimuli that were also either up or down and large or small, but were right or left and large or small for Cond. 2; (*c*) right or left stimuli that were also either up or down and bright or dim, but were right or left and up or down for Cond. 2; (*d*) right or left stimuli that were also either up or down and bright or dim, but were right or left and bright or dim for Cond. 2; (*e*) right or left stimuli that were also either large or small and bright or dim, but were right or left and large or small for Cond. 2; and (*f*) right or left stimuli that were also either large or small and bright or dim, but were right or left and bright or dim for Cond. 2. Condition 4 for Group I (scale value of stimulus complexity = 4 bits) consisted of targets that were right or left, up or down, large or small, and bright or dim.

For Group II, Cond. 1 consisted of just two targets, one up and the other down, constant with respect to right-left position, size, and brightness. The separation of the two stimuli was 1 in. The same kinds of dimensional combinations were followed as in Group I in that Cond. 2 consisted of three patterns representing the combinations of the up or down variation with the right or left, large or small, or bright or dim dimensions. There were also six patterns for Cond. 3, and the final combination of all dimensions for Cond. 4.

The stimulus conditions for Groups III and IV were similarly composed. For Group III, Cond. 1 consisted of two targets that varied only in size. The large target was .34 in. in diameter, and the small one was .25 in. in diameter. For

Group IV, Cond. 1 consisted of two targets, both .25 in. in diameter, that differed in brightness by 15.17 mL. (30.57 mL. vs. 15.40 mL.).

The spatial separations for the right-left and up-down targets, the size differential for the large-small targets, and the brightness difference for the bright-dim targets conform to the requirement that these differences in the levels of the dimensions are always detectable. Also, the maximal positional variation of successive targets was 2°45′ of visual angle, so that variation in fixation from the center of the 6-in. screen should have had no systematic effect. A preliminary psychophysical experiment was conducted to determine the values of the differences that were used in the main experiment.

*Apparatus.*—The Discrimination Display Apparatus was used to present the stimuli and to record response latencies and errors. The targets were presented by rear projection of slides on a flashed glass screen. The circular screen was 6 in. in diameter and was mounted on the rear wall of a three-sided cubicle, covered on top, but open at the front. The interior of the cubicle was painted flat black, except for the gray-white screen.

A joy-stick control was mounted in front of the screen approximately 28 in. above the floor. The joy stick was 9.5 in. long. In the right-left movement plane used for the responses of this experiment, the excursion from the neutral center to the extreme right position was 23°, and to the extreme left position, 22°.

An automatic slide projector, La Belle Model 301, was used to present the slides containing the targets. Two electronic timers controlled the duration of presentation and the interval between presentations of the slides. The circuits and operation of these timers have been described previously (4, 5). Push-button start and stop controls were provided so that continuous cycles could be initiated that would present the slides for preset durations at preset intervals. The projector was mounted so that the axis of its lens system was coincident with the center of the circular screen. A Polaroid filter was mounted in front of the lens system of the projector. The brightness of the targets was changed by rotating one sheet of Polaroid, the other remaining stationary.

A Standard Electric Clock recorded the time elapsed for each trial from the onset of the target presentation to the completion of the correct response by *S*, or until the end of the fixed presentation period. An electric counter recorded the cumulative errors. A red and a green pilot light, mounted above the screen, were used to provide *S* with knowledge of his results on each trial, as indicated in the instructions to follow.

The stimuli were circular patches of light presented one at a time. Plates were drawn for each target to the same scale as required for projection. These drawings were photographed on 35-mm. film with sufficient replications so that different random orders of the slides made from the negatives could be set up in advance for direct loading into the projector magazine. The slides were selected to give minimal positional variation. Tests indicated that for a series of unselected slides, the maximum deviations of the targets from center to center were about .25 in. For the 6-in. screen, this represents about a 4% variation, which was not normally detectable in successive presentation. A calibration slide, inserted before each experimental session, permitted the alignment of the projector and screen, so that the size and location of the stimuli on the screen were constant for all sessions.

*Procedure.*—The *S* sat facing the rear wall of the cubicle on an adjustable chair which could be raised or lowered so that *S*'s eye level was the same as that of the screen. Only the interior of the cubicle was within *S*'s normal visual field, the projection and control apparatus being behind the cubicle. When the chair was placed in position in front of the screen, the distance from the screen to *S*'s eye was approximately 28 in. and the joy stick was between *S*'s knees.

A trial consisted of the presentation of a stimulus for a period of 1 sec. Since a correction method was used, the green light could also be lighted any time after the incorrect response was made, and both the red and green lights remained lighted until the return of the selector switch to the neutral position. The interval between stimulus presentations was fixed at 12 sec. except for rest periods of 2 min. between blocks of 24 trials.

The room illumination was provided by four standard fixtures with frosted globes enclosing 300-w. bulbs. The brightness of the screen as measured by a Macbeth illuminometer was 5.89 mL. The screen brightness when the slide containing the stimulus target was presented was 6.06 mL. The brightness of all targets that did not involve the intensity dimension was 30.57 mL. When the intensity dimension was involved, the bright target was also 30.57 mL., but the dim one was 15.40 mL.

The *S*s of this experiment were college students who were induced to serve in the experiment by either paying them 75 cents an hour or awarding those students enrolled in the elementary psychology classes credit points toward their final grade. Only right-handed men were used. The approximate age range of these *S*s was 17 to 30 yr.

The *S* was conducted to the experimental room and asked to sit down on the adjustable chair placed before the joy stick and screen. If

*S* was assigned to Group I, the following instructions were given:

"In this experiment, I want to find out how well you can detect differences in targets which will be flashed on the screen one at a time. The targets are simply small circular patches of light or dots. More importantly, what I want to find out is how quickly you can respond or react to the targets by moving this joy stick right or left. If the target is to the right, your job is to move the joy stick as quickly as you can to the right. If the target is left, move the joy stick to the left. If you do this properly each time a target appears, the green light will light up. If you should throw the joy stick the wrong way, the red light will light. The red light will also come on if you are too slow. Since I'm interested in how fast you can do this, I've set a fixed brief period of time after which, if you still have not made the correct movement, the red light automatically comes on. Finally, you may get a red light if you do not move the joy stick all the way to the right or left, or, if you miss the contact. If you do get the red light, keep trying to do the right thing to make the green light come on.

"The experiment will only last about 40 or 45 min. from now, and we'll stop often to rest. So I would appreciate it if you would try your best while a series of targets is being presented."

For *S*s in Groups II, III, and IV the instructions were the same except that the appropriate stimulus variations, up-down, large-small, or bright-dim, were used in describing the task required.

For the conditions in all groups corresponding to the 1-bit scale value of stimulus complexity, a single random sequence of 24 targets was constructed. The sequence was not completely random, since the restrictions were imposed that there be no more than five successive stimuli requiring the same response, and that there be an equal number of right and left responses for each block of 24 trials. The other stimulus conditions were composed of random sequences of stimuli, the same for all groups, but with the same restriction as to the runs of like responses and constructed such that each level of each dimension involved in the particular stimulus condition occurred equally often with each level of every other dimension.

## Results

The reaction time and error measures provided essentially the same results. Only the reaction time data are presented. There is evidence that the reaction time measures were more reliable for the particular conditions of the experiment.

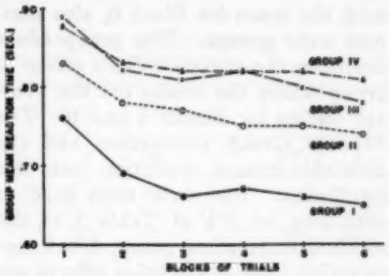


FIG. 1. Learning curves for the reaction time measure. The stimulus-response compatibility relations for the groups involve right or left responses to right or left stimuli (Group I), to up or down stimuli (Group II), to large or small stimuli (Group III), and to bright or dim stimuli (Group IV).

*Acquisition phase.*—The initial 48 trials of the experimental session constituted the acquisition phase. For each *S*, mean reaction times were computed for successive blocks of eight trials. The group mean reaction times for these units of practice are shown in Fig. 1. In every group, there is a general decrease in reaction time with successive stages of practice. The mean reaction times for Group I (right-left) are lower than for any other group. Group II (up-down) mean reaction times are lower than are those for Groups III (large-small) and IV (bright-dim), but the latter groups yield learning curves that are almost indistinguishable from each other.

An analysis of variance was carried out using the reaction time totals for the first and last blocks of learning trials. A summary of this analysis is presented in Table 1. The main effects of blocks and groups are significant at the .5% level of confidence. The significant block effect is a comparison of the mean reaction time, summed over groups, for Block 1

with the mean for Block 6, also summed over groups. The group effect describes the variation from group to group where the means are the average values for Blocks 1 and 6. The Block × Group interaction and the *Ss*-within-groups variation are not significant. The error term used in obtaining all *F*'s of Table 1 is the residual error after group, block, interaction, and *S* variation effects are removed.

*Transfer phase.*—After the initial learning trials, each *S* received 24 trials of each of the four stimulus conditions representing the four degrees of stimulus complexity as previously defined. Figure 2 is a plot of the mean reaction time obtained during each condition for each group. In general, there is an increase in reaction time with increases in stimulus complexity as defined by the experimental stimulus conditions. The over-all reaction time level increases progressively from Group I to Group IV.

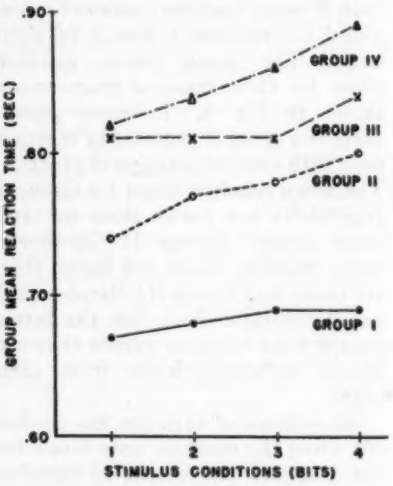


FIG. 2. The stimulus complexity functions for the reaction time measure. The ordinate is the mean reaction time per *S* per trial for the blocks of 24 trials of each stimulus condition. The stimulus conditions are represented by successive scale values of stimulus complexity on the abscissa. The parameter is stimulus-response compatibility.

In order to determine the statistical significance of these results, separate analyses of variance were performed for each group. The procedure for analysis is that described by Grant (3). Table 2 presents a summary of the results of these analyses. In every case, stimulus conditions and *Ss* within sequences are significant. The error mean square was used to test condition, ordinal-position, *Ss*-within-sequences, and square-uniqueness effects. The *F* ratio for the sequence effect was obtained by dividing the sequence mean square by the *Ss*-within-sequences mean square. The apparent significance of sequences, if the pure error mean square is used as the divisor for the *F* ratio, can thus be attributed to consistent differences between *Ss* of the different sequences. It is not possible to determine the effects of stimulus pattern in the present design since stimulus pattern is completely confounded with *Ss*.

The breakdown of the total sum of squares for stimulus conditions into individual *df* provides information rel-

TABLE 1

SUMMARY OF THE ANALYSIS OF VARIANCE FOR THE REACTION TIME MEASURE DURING ACQUISITION

| Source | *df* | MS | *F* |
|---|---|---|---|
| Block | 1 | 29.46 | 95.03* |
| Group | 3 | 13.19 | 42.55* |
| I and II, III, IV | 1 | 32.56 | 105.03* |
| II and III, IV | 1 | 6.97 | 22.48* |
| III and IV | 1 | 0.03 | 0.01 |
| B × G | 3 | 0.32 | 1.03 |
| *Ss* within groups | 92 | 0.39 | 1.26 |
| Residual | 92 | 0.31 | |
| Total | 191 | | |

* Significant at the .5% level of confidence.

TABLE 2

SUMMARY OF THE SEPARATE LATIN-SQUARE ANALYSES OF VARIANCE OF THE REACTION TIME MEASURES DURING THE TRANSFER PHASE

| Source | *df* | Group I | | Group II | | Group III | | Group IV | |
|---|---|---|---|---|---|---|---|---|---|
| | | MS | *F* | MS | *F* | MS | *F* | MS | *F* |
| Condition | 3 | 1.68 | 4.80* | 8.72 | 17.10* | 2.99 | 6.80* | 10.98 | 24.40* |
| Linear | 1 | 4.69 | 13.40* | 25.21 | 49.43* | 6.19 | 14.07* | 25.93 | 57.62* |
| Quadratic | 1 | 0.24 | 0.69 | 0.29 | 0.57 | 2.70 | 6.14* | 0.64 | 1.42 |
| Cubic | 1 | 0.11 | 0.31 | 0.65 | 1.27 | 0.07 | 0.16 | 6.37 | 14.16* |
| Sequence | 3 | 2.45 | 0.09 | 2.49 | 0.22 | 3.88 | 0.42 | 1.17 | 0.29 |
| Ordinal Position | 3 | 0.23 | 0.66 | 1.24 | 2.43 | 0.59 | 1.34 | 0.17 | 0.38 |
| *Ss* within | 20 | 8.93 | 25.51* | 11.13 | 21.82* | 9.18 | 20.86* | 4.05 | 9.00* |
| Square Uniqueness | 6 | 0.14 | 0.40 | 0.88 | 1.73 | 0.14 | 0.32 | 0.50 | 1.11 |
| Error | 60 | 0.35 | | 0.51 | | 0.44 | | 0.45 | |

* Significant at the 5% level of confidence.

ative to the nature of the function, yielding components of the stimulus-conditions total sum of squares that are associated with linear, quadratic, or cubic variations (see Snedecor [8] for a discussion of the use of orthogonal polynomials in accomplishing this breakdown). The results indicate that the linear component is highly significant for all groups. For Groups III and IV, there is an indication of significant deviations from linearity in that the *F* for the quadratic component for Group III and the *F* for the cubic component for Group IV are significant. These results probably reflect the greater decrement in proficiency in terms of the reaction time measure for Groups III and IV for the 4-bit condition. A further indication of this effect is the Group × Conditions interaction which the combined analysis, described in the next paragraph, shows to be significant.

A combined analysis based on the reaction time data for all groups is possible and meaningful since the homogeneity of the group variances is reasonable in the light of the Bartlett test results. Table 3 presents the summary of the results of this analysis. The combined analysis is useful in that it provides a single error estimate based on all of the experimental data, and permits an evaluation of the group differences and certain interactions between groups and orthogonal variables of the latin squares. The meaning attached to several of the sources of variation in the combined analysis differs from that described in the separate analyses. The condition effect for the combined analysis represents the over-all or average effect summed over groups. It is significant at the 5% level of confidence, and only the linear component of it is significant. The over-all ordinal position and sequence effects are not significant. The group effect is significant, and the same relationship between the separate groups is indicated by the comparisons for the individual *df* as was evidenced for the learning trials. The Group × Conditions interaction is also significant and reflects the differential response of the groups to the stimulus conditions. The *Ss*-within-distinct-sequences variation is highly significant. "Distinct" is used to show that this source represents the total variation of *Ss* summing only over the stimulus

TABLE 3

SUMMARY OF THE COMBINED LATIN-SQUARE ANALYSIS FOR THE REACTION TIME MEASURES DURING THE TRANSFER PHASE

| Source | *df* | MS | *F* |
|---|---|---|---|
| Condition | 3 | 19.97 | 45.38* |
| Linear | 1 | 58.44 | 133.95* |
| Quadratic | 1 | 0.50 | 1.14 |
| Cubic | 1 | 0.10 | 0.23 |
| Ordinal Position | 3 | 0.55 | 1.25 |
| Sequence | 3 | 1.23 | 0.15 |
| Group | 3 | 296.48 | 35.63* |
| Between I and II, III, IV | 1 | 731.23 | 87.89* |
| Between II and III, IV | 1 | 126.19 | 15.17* |
| Between III and IV | 1 | 31.99 | 3.84 |
| G × C | 9 | 1.46 | 3.32* |
| G × OP | 9 | 0.52 | 1.18 |
| G × S | 9 | 2.92 | 0.35 |
| *Ss* within distinct sequences | 80 | 8.32 | 18.91* |
| Square uniqueness | 6 | 0.42 | 0.95 |
| G × SU | 18 | 0.47 | 1.06 |
| Pooled latin-square error | 240 | 0.44 | |
| Total | 383 | | |

* Significant at the 5% level of confidence.

conditions. It includes within its total sum of squares an *Ss*-within-sequences variation for pooled *Ss* of the same sequence within the latin square. No other significant effects were found.

## DISCUSSION

The reaction time measures for the initial 48 trials of the experimental session describe four learning curves, one for each of the experimental groups. Although the actual stimuli to which *Ss* of the four groups responded were different, the stimulus complexity of the stimulus conditions is the same and is equal to a scale value of 1 bit. All groups show significant improvement. There is no evidence that the rate or extent of improvement is different for the different groups. However, the over-all reaction time level does vary from group to group reflecting the operation of the variable that has been defined as stimulus-response compatibility.

Group I has the most compatible stimulus-response relation. The right control movement to a target appearing to the right of the display appears to conform to a very general "population stereotype" (1). The other groups exhibit less compatible stimulus-response relations. However, the tasks for both Groups I and II are similar in that both require a spatial displacement of the control for a spatial displacement of the target. The tasks for Groups III and IV do not involve the same spatial to spatial characteristic and for this reason are presumed to provide less compatible relationships. The group differences remain fairly constant throughout the course of learning.

The transfer phase provides a stimulus complexity function that is essentially free from learning effects. This function shows that the latency of the discriminative response increases with increasing stimulus complexity values. The same ordering of the groups with respect to the stimulus-response compatibility dimension is observed in the transfer phase as is found in the acquisition phase, but the shape of the stimulus complexity functions are different for the groups. These findings suggest that proper application of the stimulus complexity concept requires specification of the stimulus-response compatibility relationship. As a parameter of stimulus complexity, stimulus-response compatibility appears not only to affect the level of response proficiency, but also the shape of the stimulus complexity function.

The fact that the proficiency measures obtained in the transfer phase were substantially influenced by the changes in the stimulus conditions, even though a relatively stable performance level had been established in the acquisition phase, suggests that the concept of stimulus complexity may be valuable in specifying the equivalence of discriminative tasks. Training for tasks that involve perceptual-motor discrimination skills can be carried out efficiently only if the functional relationships among proficiency measures and the variables governing the acquisition of such skills are

known. Description, prediction, and evaluation of such relationships may be facilitated by the use of the stimulus complexity dimension. If, for example, a variable such as the amount of practice or distribution of practice must be explored in order to determine the optimum training conditions for a class of perceptual-motor tasks, quantitative specification of the tasks could be carried out in terms of the stimulus complexity values, even though the discriminanda vary over many dimensions. Relating the acquisition variable to the stimulus complexity values would provide the basis for direct comparisons among tasks with the same complexity value. In addition, a criterion for predicting the training conditions necessary to produce the desired proficiency levels for tasks of different complexity would be established.

## Summary

A technique for quantifying stimulus aspects of discrimination tasks was presented. Visual stimuli were used that varied in horizontal spatial separation, vertical spatial separation, size, and brightness. The motor responses were right or left movements of a joy-stick control. Four groups of 24 *S*s each, differentiated in terms of a variable referred to as stimulus-response compatibility, received 48 acquisition trials. The stimulus conditions for these trials represented the least complex discrimination corresponding to a scale value of stimulus complexity of 1 bit. Following the acquisition trials, *S*s of each group received four blocks of 24 trials each. These blocks represented four levels of stimulus complexity corresponding to the scale values of 1, 2, 3, and 4 bits. The purpose of this phase was to determine the relationship between stimulus complexity and the proficiency measures, reaction time and errors, of the discrimination performance.

The results of the analyses of the reaction time data showed that reaction time during acquisition decreased at approximately the same rate and to the same extent for all groups. An asymptotic reaction time level was reached within the 48 acquisition trials. The over-all group reaction time levels were different. The group in which the horizontal spatial discrimination was associated with the right-left movements showed the lowest over-all reaction time. Next was the group in which the vertical spatial discrimination was associated with the right-left movements. The other two groups, in which the size and brightness discriminations were associated with the right-left movements, had still higher mean reaction times, but did not differ significantly between themselves. Reaction time was found to increase as the scale values of stimulus complexity increased. The relationship was essentially linear, but the exact nature of the function was somewhat different for the different groups. Analyses of the error data yielded similar results.

## References

1. Fitts, P. M. Engineering psychology and equipment design. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 1287–1340.
2. Garner, W. R., & Hake, H. W. The amount of information in absolute judgments. *Psychol. Rev.*, 1951, **58**, 446–459.
3. Grant, D. A. The latin square principle in the design and analysis of psychological experiments. *Psychol. Bull.*, 1948, **45**, 427–442.
4. Gregg, L. W. Fractionation of temporal intervals. *J. exp. Psychol.*, 1951, **42**, 307–312.
5. Hunter, T. A., & Brown, J. S. A decade-type electronic interval timer. *Amer. J. Psychol.*, 1949, **62**, 570–575.
6. Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.
7. Shannon, C. E. A mathematical theory of communication. *Bell Syst. Tech. J.*, 1948, **27**, 379–423, 623–656.
8. Snedecor, G. W. *Statistical methods*. Ames, Iowa: Collegiate Press, 1946.

(Received for early publication
April 1, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# EXTRA-TASK PERFORMANCE AS A MEASURE OF LEARNING A PRIMARY TASK [1,2]

HARRY P. BAHRICK

*Ohio Wesleyan University*

MERRILL NOBLE AND PAUL M. FITTS

*The Ohio State University*

The distinction between the basic process of learning and the manifestations of this basic process is often explicitly or implicitly drawn. Thus, Hull (3) distinguishes between habit strength on the one hand and amplitude, latency, probability of occurrence, and resistance to extinction on the other. Stevens (6) distinguishes between *measures* of psychological processes and *indicants* of these processes. Empirical confirmation of the cogency of this distinction is to be found in the low correlations among various measures of (presumably) the same learning process (2, p. 138) and in those instances in which differential levels of learning have been detected by some measures and not by others (5, ch. 8). However, realization of this distinction has not generated much systematic research oriented either toward the identification of differences among the most commonly used performance measures, or toward the development of new techniques for measuring aspects of learning not revealed by present performance measures.

The present study was undertaken to determine whether performance on a secondary task, practiced simultaneously with a primary task, can serve to reveal learning on the primary task that is not revealed by conventional measures on the primary task itself. The study was based on the widely held notion that with continued practice tasks become less susceptible to interference. It was hypothesized that reduced susceptibility to interference is characteristic only of tasks that possess some degree of repetitiveness.[3] It was reasoned that proprioceptive and tactual stimulation associated with one motor response may come to elicit the next response when repetitive sequences are learned. This kind of motor-sequence learning should permit *S* to respond more effectively to other concurrent stimulation and should therefore be revealed in improved extra-task performance. The learning of response sequences need not necessarily affect the quality of individual motor responses, and common

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-10528 with The Ohio State University Research Foundation, monitored by the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole or in part by or for the United States Government. The experiment was reported at the 1953 meeting of the American Psychological Association under the title "The Effect of Motor Task Redundancy and Level of Practice upon Interference between Two Tasks."

[2] The authors are indebted to Dr. David Bakan for many constructive criticisms.

[3] Repetitiveness is conceived as a continuum, with a fixed order of occurrence (here called repetitive) at one extreme and a random sequence with a rectangular distribution of the stimuli at the other. The several restrictions placed on the generation of the sequence called "random" in this paper place it at some distance from the end of the continuum.

performance indices may therefore fail to reveal it.

To test this hypothesis, *Ss* were trained either in a repetitive or in a random version of a motor task. The two versions were adjusted in such a way that a primary task performance index of the various groups remained at a comparable level at each stage of training. Concurrent performance of an extra task was required either early or late in practice on both versions of the motor task.

## Method

### *Apparatus*

A schematic drawing of the apparatus is shown in Fig. 1. It consisted of a vertically rotating drum of 48-in. circumference driven by a synchronous motor. The surface of the drum was of translucent plastic. During the experiment this surface was covered by one of two interchangeable black paper masks in which were punched ½-in. holes. The holes were arranged in a prearranged sequence along five equally spaced vertical columns. A light source was located inside the drum. As the drum was rotated, the ½-in. holes appeared as dots of light which moved downward along the five columns. The speed of rotation of the drum was varied by changing the size of a driving wheel on the synchronous motor.

The dots of light were viewed through a curved translucent viewing screen on which was drawn a horizontal crossline. No two dots reached this line simultaneously. Five finger keys were located below the screen and corresponded in position to the five light columns. If a corresponding key was pressed during the transit of a light dot behind the crossline, one of five hit counters located at *E*'s control panel was actuated. Other impulse counters recorded the total number of dots (stimuli) that passed the crossline and the number of key-pressing responses. A small knowledge-of-results light located directly above the screen flashed whenever a hit was scored.

### *Subjects*

Seventy male and 70 female right-handed volunteer students from elementary psychology courses at The Ohio State University were divided by means of a table of random numbers into four groups of 35 *Ss* each, with approximately equal proportions of men and women in each group.

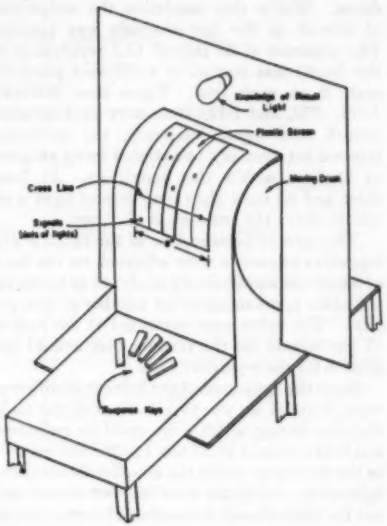


Fig. 1. A schematic drawing of the apparatus

### *Experimental Tasks*

The primary task employed the visual-motor apparatus just described. The secondary task consisted of mental arithmetic.

*Visual-motor task.*—The *Ss* were instructed to place the fingers of their right hand lightly on the five keys and to press the corresponding key at the instant of transit of a light dot behind the crossline. A trial consisted of a sequence of 60 dot transits.

Random and repetitive versions of the primary task were obtained by varying both the temporal and the spatial sequence of the dots. For the repetitive version dots were arranged in an irregular temporal and spatial sequence that was repeated exactly after every fifth stimulus. If we number the columns of lights from left to right, the spatial pattern of the repetitive sequence was 1, 5, 3, 4, 2; 1, 5, 3, 4, 2; etc. The temporal sequence, beginning with the interval between lights 1 and 5 in the spatial sequence, was 1.125, .750, .750, .375, .750 sec. Either five or six dots of light were visible above the crossline at all times, depending on the position of the drum.

For the random version of the task, dots were arranged so that 10 stimuli occurred in each of the five channels during one revolution of the

drum. Within this restriction the assignment of stimuli to the five channels was random. The sequence of 60 stimuli (1.2 revolutions of the drum) was started at a different point for each *S* on each trial. Three time intervals (.375, .750, and 1.125 sec.) were used between stimuli with equal frequency, the particular interval between any two stimuli being assigned at random within this restriction. At least three and no more than nine dots of light were visible above the crossline at all times.

The rates of presentation of the random and repetitive sequences were adjusted, on the basis of data from a preliminary study, so as to equate the tasks approximately for number of hits per trial. The lights were presented at the rate of 51 per minute for the random task and 81 per minute for the repetitive task.

Since the lights passed the hairline at different rates of speed for the two versions of the task, the time during which a hit could be registered was held constant at .13 sec. ($\pm$.065 sec. relative to the instant at which the crossline bisected the light spot). Only the time between stimuli and not the time allowed for scoring a hit was varied.

*Mental subtraction task.*—In the secondary task *S* was asked to report the difference between two previously read numbers. The *E* began by reading two numbers. As soon as *S* gave the answer, *E* read another number and *S* determined the difference between it and the preceding number read by *E*. The *E* recorded each answer given and immediately read out a new number regardless of whether the answer was right or wrong. If *S* did not respond within 5 sec., another number was read and the test continued. The duration of a trial on the arithmetic task was the same as for the visual-motor task. The score was the number of correct responses per minute.

### *Procedure*

The four experimental groups were: (*a*) repetitive, low practice, (*b*) repetitive, high practice, (*c*) random, low practice, and (*d*) random, high practice. The two high-practice groups were first given 15 trials on the visual-motor task alone and then three trials during which the motor and the arithmetic tasks were performed concurrently. The two low-practice groups received only three trials on the motor task prior to practice on both tasks concurrently.

All *S*s were given the same instructions. After each trial *S*s were told the number of hits they had scored in the visual-motor task and, after appropriate trials, the number of correct arithmetic responses. Rest pauses of 20 sec. were given between trials, and a 3-min. rest was given to the two high-practice groups after every fifth trial.

Instructions for the arithmetic task and six practice problems were given just prior to the three trials of concurrent practice on both tasks.

## Results

The means and *SD*'s for both tasks for all four groups are given in Table 1. An *F* test of the variations among the four means on the arithmetic tasks was found to be significant at the .01 level ($F = 5.79$). An *F* test of the variation in mean performance on the motor task during the first three trials was not significant ($F = 1.01$). All other comparisons between groups are based upon simple *t* tests.

The following are the principal results:

*a.* The two low-practice groups did not differ significantly from each other on either the motor or the arithmetic task ($p > .05$ in each instance). These groups were comparable at the outset of training on the visual-motor task, they suffered comparable decrements on the visual-motor task when the arithmetic task was introduced, and their performance on the arithmetic task was comparable.

*b.* The two high-practice groups differed significantly only in their arithmetic performance ($p < .01$). The performance on the motor task was comparable for the two groups at the outset of training, showed similar improvement with training, and dropped by comparable amounts when the arithmetic task was added ($p > .05$ in each instance).

*c.* The mean arithmetic score of the repetitive high-practice group was significantly higher than that of the repetitive low-practice group ($p < .01$). However, the mean arithmetic score of the random high-practice group was not significantly higher

TABLE 1

MEANS AND *SD*'S FOR THE MOTOR AND ARITHMETIC TASKS DURING CRITICAL BLOCKS OF TRIALS
($N$ = 35 in each group)

| Trials | High Practice | | | | Low Practice | | | |
|---|---|---|---|---|---|---|---|---|
| | Random | | Repetitive | | Random | | Repetitive | |
| | Mean | *SD* | Mean | *SD* | Mean | *SD* | Mean | *SD* |
| Motor Task | | | | | | | | |
| 1–3 | 28.68 | 9.46 | 24.70 | 11.02 | 28.27 | 9.70 | 24.89 | 11.33 |
| 4–6 | | | | | 22.29 | 9.33 | 18.11 | 13.88 |
| 13–15 | 37.79 | 11.84 | 33.36 | 13.15 | | | | |
| 16–18 | 26.91 | 10.55 | 23.77 | 12.74 | | | | |
| Arithmetic Task | | | | | | | | |
| 4–6 | | | | | 9.43 | 2.62 | 9.51 | 3.66 |
| 16–18 | 9.97 | 4.03 | 12.84 | 5.01 | | | | |

than that of the random low-practice group ($p > .05$).

## DISCUSSION

It is evident that extra-task (arithmetic) performance late in practice was superior to extra-task performance early in practice when the sequence of stimuli in the concurrent motor task was repetitive. No such difference occurred when the motor task was random. Since scores on the motor task itself remained comparable at different levels of training for the random and repetitive groups, these results support the hypothesis that extra-task performance can be used to measure certain kinds of improvement on a primary task which are not evident from conventional scores on that task itself. The results also support the hypothesis that, for the level of skill attained, the improvement shown by this measure is peculiar to a repetitive motor task which permits the learning of fixed response sequences.

At least one interesting question regarding these results remains unanswered. The learning effect shown by the repetitive high-practice group was manifested in improved performance on the arithmetic task. Learning might equally well have led to reduced interference on the motor task. A possible explanation for the direction of the effect can be found in an earlier study (1) where it was shown that the one of two tasks which is perceived as more important is performed with little decrement, especially under conditions of high motivation. If we assume that in the present experiment the continuous visual-motor task was perceived as the more important task, it would follow that most of the differential interference effects would be manifested on the arithmetic task, while the performance decrement on the visual-motor task would be small and at a comparable level for all the groups. This interpretation must remain tentative until it is substantiated by further data.

The present experiment was terminated at a level of training where primary-task scores had begun to asymptote. It remains to be determined whether extra-task scores may continue to improve even after primary-task scores fail to show further improvement. The aspect of learning which makes performance on one task more compatible with another concurrent activity may be particularly characteristic of the later stages of learning of certain tasks. If this should be the case, extra-task performance may provide a useful way

of measuring overlearning under conditions of extended practice. Since the writings of William James (4, ch. 4), authors have casually spoken of certain activities becoming automatized with much overlearning, but little is known about the operations that would define this process or variables that permit its occurrence. The present investigation indicates that it may be possible to develop quantitative measures of both the antecedent and consequent conditions of such a process.

## Summary

The hypothesis was advanced that the aspect of learning which makes performance of a task more compatible with other concurrent activities is not adequately reflected by any of the conventional measures of learning. This aspect of learning, it is proposed, is made possible in part by the degree of repetitiveness of the task, and can be measured by requiring simultaneous performance of a second task at various levels of practice on the primary task.

The *S*s were trained in either a random or a repetitive version of a visual-motor task which required anticipation and precise timing. A mental arithmetic task was added either early or late in practice on the visual-motor task.

Arithmetic performance was comparable for the random and repetitive groups if the secondary task occurred early in practice. However, arithmetic scores were significantly superior for the repetitive group when arithmetic was added late in practice. Scores on the motor task remained comparable whether the secondary task was added early or late.

These results indicate that extra-task performance on a mental arithmetic task can be used to measure an aspect of learning of a visual-motor task which is not reflected by the scores on the motor task itself, and that this aspect of learning is dependent upon the repetitiveness of the task.

## References

1. Bahrick, H. P., Fitts, P. M., & Rankin, R. E. Effect of incentives upon reactions to peripheral stimuli. *J. exp. Psychol.*, 1952, **44**, 400–406.
2. Hilgard, E. R., & Marquis, D. G. *Conditioning and learning.* New York: D. Appleton-Century, 1940.
3. Hull, C. L. *Principles of behavior.* New York: D. Appleton-Century, 1943.
4. James, W. *The principles of psychology.* New York: Holt, 1890.
5. McGeoch, J. A. *The psychology of human learning.* New York: Longmans, Green, 1942.
6. Stevens, S. S. Mathematics, measurement, and psychophysics. In S. S. Stevens, (Ed.), *Handbook of experimental psychology.* New York: Wiley, 1951. Pp. 1–49.

(Received for early publication July 6, 1954)

*Journal of Experimental Psychology*
Vol. 48, No. 4, 1954

# THE LEARNING OF SEQUENTIAL DEPENDENCIES[1]

WILLIAM F. BENNETT, PAUL M. FITTS, AND MERRILL NOBLE

*The Ohio State University*

A distinction can be made between those learning situations in which events occur in fixed sequences and those situations in which events occur in sequences governed only by statistical laws. In the latter case a further distinction can be made between the learning of probabilities governing independent events and the learning of probabilities governing ergodic sequences, i.e., sequences in which the probability of occurrence of any given event is dependent in part on preceding events in a stable time series. The present investigation is concerned with the latter type of learning situation, which will be referred to as sequential probability learning.

The acquisition of many verbal and motor skills involves the learning of sequential probabilities. For example, it has been estimated that the English language is at least 50% redundant (11). Not only are different language symbols used with unequal frequency but they are also used in characteristic sequential patterns. The development of language skills undoubtedly involves learning about these sources of redundancy. As an illustration, recent experiments indicate that recall of verbal material is predictable in part from a consideration of contextual constraints (1, 7), and it is well known that words heard in the context of a sentence are more likely to be heard correctly than words heard in isolation. Similarly, many skilled motor activities, such as piloting an aircraft, involve the learning of characteristic stimulus sequences and characteristic sequences of movements.

The sequential dependencies characteristic of continuous or discrete time series can be specified in several ways. The statistical nature of the fluctuations of a continuous ergodic signal is specified in part by an autocorrelation function. The characteristics of an ergodic sequence of discrete events in which the number of events and the extent of intersymbol influence are limited are specified completely by means of an appropriate table of transition probabilities. In both cases the magnitude of the probability constraints can be expressed as the relative redundancy of the sequence, i.e., 1 − the ratio of actual to potential information.[2]

In the present experiments *S*s tried to predict the successive events in sequences which had been generated by selecting stimuli at random in accordance with a table of transition probabilities. In order to isolate the learning of sequential dependencies, as contrasted with the learning of the probabilities associated with independent events, first-order probabilities were held constant. Thus, each stimulus occurred equally often in the

[1] This research was supported in part by the United States Air Force under Contract No. AF 33(038)-10528 with The Ohio State University Research Foundation, monitored by the Skill Components Research Laboratory of the Air Force Personnel and Training Research Center. Permission is granted for reproduction, translation, publication, use, and disposal in whole and in part by or for the United States Government. Part of the data reported was collected as part of an M.A. thesis by the first author.

[2] The measure of relative redundancy has been described fully by Miller and Frick (6). Descriptions of the measures of uncertainty and behavioral stereotypy, referred to later in this report, also are described by these authors.

sequences studied, but stimuli occurred only in certain sequences.

A series of four learning experiments was carried out to answer two questions. The first question was: How difficult is it to learn to respond effectively to unequal digram probabilities as contrasted with unequal trigram probabilities?[2] In the case of unequal digram probabilities, an optimum prediction of the next event can be made on the basis of a single preceding event. In the case of equal digram but unequal trigram probabilities an optimum prediction can be made only if two preceding events are considered.

Most *S*s do not use all possible response sequences equally often but tend to use "preferred" patterns of responses. The second question studied was: What is the effect of sequential guessing habits on the ease of learning sequential dependencies?

## Experiment I

This experiment was designed to determine sequential guessing habits in a 5-alternative choice situation in which there were equal frequencies of occurrence of all engrams,[3] digrams, and trigrams in a random sequence, i.e., in a situation in which it is theoretically impossible to achieve, on the average, more than a chance number of successes. Empirical tables of transition probabilities specifying such guessing habits were needed in order to generate the sequential patterns required in subsequent experiments.

[3] *Engram*, *digram*, and *trigram* are terms used to denote the extent of interevent influence. Thus, *engram* refers to a sequence in which an event is completely independent of previous events, *digram* to a sequence in which the probability of occurrence of an event is dependent in part upon the immediately preceding event, and *trigram* to a sequence in which the probability of occurrence of an event is dependent in part upon the two preceding events.

### *Method*

*Apparatus.*—The apparatus used throughout the four experiments reported here provided a task in which *S* predicted the sequential occurrence of five lights. The *S*'s display and control panel is schematically illustrated in Fig. 1. It consisted of (*a*) five push-button switches, (*b*) five blue response lights, and (*c*) five red knowledge-of-results lights. Each pair of red and blue lights corresponded to the response button directly below it. The blue response lights were controlled by the corresponding push-button switches, and the red knowledge-of-results lights by *E*.

On *E*'s panel (not shown) were five lights that indicated which of the response buttons had been selected by *S*, and five switches which permitted *E* to activate the red knowledge-of-results lights on *S*'s panel. A sixth switch turned off all the lights.

*Subjects.*—Twenty undergraduates, 10 men and 10 women, enrolled in elementary psychology courses at The Ohio State University served as *S*s.[4]

*Pattern of correct reponses.*—An arbitrary sequence of correct responses was constructed by following a restricted random procedure which insured that each engram, each digram, and each trigram occurred with equal probability in the complete series. A sequence of 125 symbols placed end to end is the minimum length that will fulfill the above requirements in a 5-alternative situation. In the sequence employed in this study each possible engram appeared with a frequency of 25, each possible digram appeared with a frequency of 5, and each possible trigram appeared with a frequency of 1 in 125 events. No attempt was made to equalize sequential dependencies of a higher order than trigrams.

*Procedure.*—The *S*s were instructed to guess the sequence of the knowledge-of-results lights and to indicate their predictions by pushing the appropriate buttons. They were told only that the knowledge-of-results lights would follow a predetermined pattern. The instructions did not mention sequential dependencies. A 4-sec. rate of responding was established by means of a timer connected to a buzzer.

After each response *E* recorded *S*'s prediction and indicated the predetermined correct response by turning on one of the red lights. The *E* then turned off all the lights before the buzzer sounded for the next prediction. This procedure was

[4] Elementary psychology students are required to serve in experiments to fulfill part of their laboratory assignment. They are permitted to choose among various investigations being conducted; in this sense *S*s of all of the experiments reported here are volunteers.

continued until each *S* made a total of 250 responses (two repetitions of the sequence of 125 lights).

*Results*

The frequencies of occurrence of each engram, digram, and trigram in the response sequences were tabulated and the results for all *S*s pooled. Only *S*s' response sequences were analyzed; the sequences of the knowledge-of-results lights were not considered. Chi-square tests indicated that both the digram and trigram frequencies of the pooled responses of all *S*s differed from chance distributions beyond the .01 level of confidence, and that the engram distribution differed from chance at approximately the .06 level of confidence. Thus, as found by others (**2, 3, 9, 10**), the *S*s' responses were more structured than was the stimulus series itself.

Newman and Collier (**9**) and Chapanis (**2**) have reported that *S*s tend to avoid repetitive sequences such as 5-5-4, 1-2-2, and 1-5-1. This tendency was also found with the present *S*s. Of the five repetitive digrams (1-1, 2-2, etc.), four occurred with frequencies less than expected by chance, but for only one, the sequence 3-3, was the reduction below a chance level statistically significant ($\chi^2 = 5.77$; $p < .02$). Forty-six of the 65 repetitive trigrams occurred with less than the expected frequency. Eight of these occurred with a frequency reliably below the expected value as shown by $\chi^2$ ($p < .05$), while only two repetitive trigrams (2-2-2, and 5-5-1) occurred with a frequency greater than expected at the .05 level of confidence. Twenty of the repetitive trigrams do not contain a repetitive digram; of these 18 were used with less than the expected frequency, but only four of them, 1-5-1, 3-5-3, 5-1-5, and 5-3-5, differed from chance at or below the .05 level of confidence as shown by $\chi^2$. In general, therefore, the tendency to avoid repetitive sequence is not universal even though it appears to be consistent.

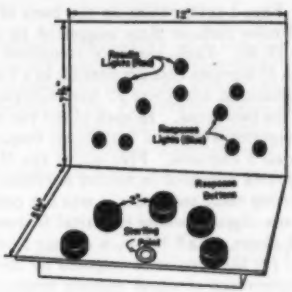

FIG. 1. Apparatus used in the four experiments

## EXPERIMENT II

This experiment was designed to determine (*a*) the difficulty of learning sequential patterns characterized by equal engram but unequal digram and trigram probabilities, and (*b*) the persistence, in terms of number of trials, of the influence of guessing sequence habits. Separate groups of *S*s learned one of two sequences. Within the restrictions imposed by the method employed for selecting successive stimuli, one sequence was made to agree as closely as possible with empirically determined guessing habits (concordant sequence) and one was made as dissimilar as possible to guessing habits (discordant sequence).

*Method*

*Subjects.*—Forty men and women volunteers from elementary psychology courses at The Ohio State University were assigned alternately to one of two groups of 10 men and 10 women per group. None of the *S*s had served in the preceding experiment.

*Sequence of correct responses.*—The concordant and discordant sequences were constructed on the basis of the empirical digram frequency

data of Exp. I rather than on the basis of generalized rules such as those suggested by other writers (2, 9). Each sequence contained only 10 of the 25 digrams that are possible in a 5-alternative situation, and these 10 were different for each of the two series. In each of the two series the 10 digrams were used with equal frequency, as were all 5 engrams. Five out of the 25 possible digrams appeared in neither sequence. In constructing these sequences it was not possible to find ten digrams whose empirical frequencies were all above, or all below, a chance level because of (*a*) the restriction imposed by the further requirement of equal engram frequencies, and (*b*) the fact that guessing habits were not strongly marked. Only half of the digrams used to make up the concordant and discordant sequences were ones that had been found in Exp. I to differ significantly from a chance value. (The median relative frequency of the digrams making up the concordant sequence was .042, and of the digrams in the discordant sequence, .037. The relative frequency expected by chance was .040.)

The two sequences were 250 events in length and were so constructed that when each sequence was placed end to end so as to form a closed cycle, each of the five engrams occurred 50 times and was followed equally often by one of two alternatives. For example, in the concordant sequence Light 1 was followed 50% of the time by itself and 50% of the time by Light 4. Within these limitations each engram and each digram occurred at random. Twenty different trigrams also were inherent in the sequence. Each of these occurred from 12 to 14 times in the total sequence of 250 stimuli. The digram sequential dependencies which define the two patterns of correct responses are shown schematically in Fig. 2.

*Procedure.*—The instructions and procedure were the same as in Exp. I. The additional precaution was taken of starting each *S* at a different place in the sequence.

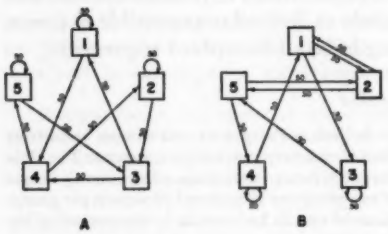


FIG. 2. Schematic representation of the digram patterns: 2A shows the concordant sequence; 2B shows the discordant sequence

*Results*

Responses were scored as appropriate when *S* predicted either of the two possible correct alternatives. Thus, it was possible for *S* to earn a score of 100% appropriate responses, even though on the average it was possible for him to predict the next light correctly only 50% of the time. Performance curves for the two groups, showing the mean number of appropriate responses per block of 25, are given in Fig. 3.

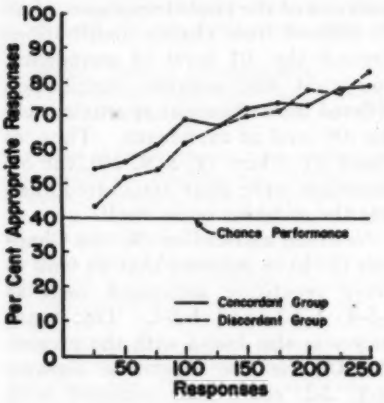


FIG. 3. Percentage of appropriate responses with the digram patterns of Exp. II. Lines on the abscissa represent successive blocks of 25 responses, i.e., responses 1–25, 26–50, etc. Lines on the ordinate refer to the proportion of appropriate responses relative to total responses made. An appropriate response is either of the two equally probable events.

Both the Concordant and Discordant groups showed a steady increase in the mean number of appropriate responses per block of 25 predictions. The effect of guessing habits is evidenced by the tendency of the Concordant group to make a greater than chance number of appropriate predictions on the initial block of trials (the mean percentage appropriate responses was 54.2 as compared to a

chance frequence of 40.0%). The Concordant group was superior to the Discordant group on 8 of the 10 blocks of responses. However, only the difference between the first of the ten separate blocks of responses is significant ($t = 3.80$; $df = 38$). When the data for the first four blocks of trials are pooled, the difference is significant between the .02 and .01 levels of confidence ($t = 2.66$; $df = 38$) in favor of the Concordant group. Performance during the first 25 responses was approximately at a random level (40% appropriate responses) for the Discordant group. Both groups were making approximately 80% appropriate responses by the end of the sequence of 250 events.

Table 1 shows the index of behavioral stereotypy, $C_s$ (6), for both groups for the first and final blocks of 25 responses, considering (*a*) single responses, (*b*) pairs of responses, and (*c*) series of three responses. The measure $C_s$ is defined as 1 − the relative uncertainty per response. The data of Table 1 indicate that behavioral sequences tend to become somewhat more stereotyped as practice is continued.

TABLE 1

BEHAVIORAL STEREOTYPY ($C_s$) FOR THE FIRST AND LAST BLOCK OF 25 RESPONSES CONSIDERING NO PRECEDING RESPONSES, THE IMMEDIATELY PRECEDING RESPONSE, AND THE TWO IMMEDIATELY PRECEDING RESPONSES

| Condition | Value of $C_s$ | | |
|---|---|---|---|
| | Length of Response Sequence | | |
| | One | Two | Three |
| Concordant group | | | |
| Responses 1–25 | .003 | .021 | .074 |
| Responses 225–250 | .001 | .035 | .099 |
| Discordant group | | | |
| Responses 1–25 | .013 | .038 | .096 |
| Responses 225–250 | .007 | .114 | .292 |

Theoretically, if the data on guessing habits derived from Exp. I are reliable when applied to a different group of *S*s, initial performance of the Discordant group should have been as much below the chance level as the initial performance of the Concordant group was above the chance level. Inspection of the performance curves of Fig. 2 indicates instead that for the first block of 25 trials both groups were equal to or above a chance level. A possible explanation for the initial high performance of the Discordant group is that sequential guessing habits were modified early in the course of the first 25 trials as a result of the fact that the sequence of reinforcements was opposed to these habits. Although sufficient data are not contained within the first block of 25 predictions to permit more than a gross analysis of within-series changes in response patterns, no evidence was found to confirm this possibility.

A more likely explanation for the failure of the Discordant group to be initially as much below chance performance as the Concordant group was above chance is unreliability of sequential guessing habit data. Sequential guessing habits represent at most a small departure from chance values, and are subject to large individual differences. Individual sequential patterns also change to some extent at different stages of practice. It is tentatively concluded, therefore, that the guessing habits of the individuals assigned to the Discordant group were such that for them the sequence was essentially a neutral sequence.

The effects of sequential guessing habits might have been more marked if *E*s, in presenting the two contrasting knowledge-of-results sequences, had worked from a table of transition probabilities which included the effects of the preceding knowledge-of-results lights rather than reflecting only the sequential effects of *S*'s own responses. In order to explore this possibility the analysis of the data of Exp. I was extended and the

TABLE 2

BEHAVIORAL STEREOTYPY ($C_s$) OF THE RESPONSES OF EXP. I AS A FUNCTION OF CONSIDERING PREVIOUS RESPONSES AND PREVIOUS KNOWLEDGE-OF-RESULTS LIGHTS

| Sequence | Value of $C_s$ | |
|---|---|---|
| | Number of Previous Events Considered* | |
| | One | Two |
| Knowledge-of-results lights | .012 | .021 |
| Ss' responses | .007 | .010 |

* When neither previous knowledge-of-results lights nor previous responses are considered, $C_s = .001$.

index of behavioral stereotypy, $C_s$, was employed for comparing the relative sequential influence on S's responses of (*a*) the sequence of knowledge-of-results lights versus (*b*) S's own preceding responses. The results of this analysis indicate that responses can be predicted with greater success when the preceding knowledge-of-results lights are used as a basis for prediction than when only the sequence of S's preceding responses is used. This finding is in agreement with the results of Hake and Hyman (3). The results of the analysis of the present data, summarized in Table 2, indicate that greater stereotypy results for a sequence of given length when previous knowledge-of-results lights are considered than when previous responses are considered.

## EXPERIMENT III

The third experiment was designed to determine the difficulty of learning a sequence in which better-than-chance predictions can be made only on the basis of two or more preceding events (trigram or higher sequences). The influence of sequence guessing habits was also investigated as in the preceding experiment, only in this case trigram guessing habits were considered.

### *Method*

*Subjects.*—Twenty men and women from elementary psychology courses at The Ohio State University were assigned alternately to two groups of two men and eight women per group. None of the Ss who had served in the preceding experiments were used.

*Sequences of correct responses.*—It is possible, when using five symbols, to generate 125 different trigrams. With the rationale used in the preceding experiment, two sequences were constructed, each containing a different two-fifths of the possible trigrams. Each sequence was 50 events in length and, when placed end to end so that the first event followed the fiftieth, all 50 trigrams occurred once each time through a cycle of 50 stimuli. In both patterns each possible engram and each possible digram in the 5-alternative situation appeared equally frequently and, within the limitations imposed by the procedure, at random in the sequence.

The two patterns were constructed in such a way that one contained as many as possible, and the other as few as possible, of the trigram response sequences favored in Exp. I. The median relative frequencies of the trigrams in the two sequences were .0092 and .0072, respectively, for the concordant and discordant sequences (probability on the basis of random responding = .0080). Fourteen of the 50 trigrams included in the concordant sequence and 10 of the 50 trigrams of the discordant sequence were significantly different from chance ($p < .05$) on the basis of $\chi^2$ tests of the data from Exp. I.

*Procedure.*—The instructions and procedure were the same as in Exp. I and II except that the cycle of 50 events was repeated ten times, and each S made 500 instead of 250 responses. Thus, each trigram appeared a total of 10 times in the sequence of 500 trials. A 5-min. rest period was given after Response 250. It should be noted, on the one hand, that the trigram sequences employed in Exp. III consisted of a repetitive cycle only 50 events in length, compared to a cycle 250 events in length in the digram sequence; on the other hand, 50 unique sequences had to be learned in this experiment as compared with 10 unique sequences in Exp. II.

### *Results*

The results of this experiment are presented graphically in Fig. 4. The Concordant group was significantly superior to the Discordant group for the total block of 500 responses ($t = 4.03$; $df = 18$). Inspection of Fig. 4 indicates that the magnitude of the

difference between the two groups remained approximately constant for about the first 400 trials.

Inspection of Fig. 4 also indicates that little learning occurred in either group. Analyses of variance indicated no tendency for the difference between the first 200 and the last 200 responses to approach statistical significance for either the Concordant or the Discordant groups. Furthermore, several *Ss* who had relatively high scores for the first 100 responses showed a decrement rather than an improvement during later portions of the series.

Although there was little learning of either of the trigram sequences, it is nevertheless possible that changes in *Ss'* predictive behavior took place during practice. This possibility was explored, using the index of behavioral stereotypy at different stages of practice as an index of a change in the response sequence. No consistent trends in the index of behavioral stereotypy were found when the first, middle, and final blocks of 50 responses of the two groups were examined, using either digrams or trigrams as a basis for prediction. For the former, $C_s$ remained at approximately .025 for each group at each stage of practice; for the latter, $C_s$ remained at approximately .070 for each group at each stage of practice.

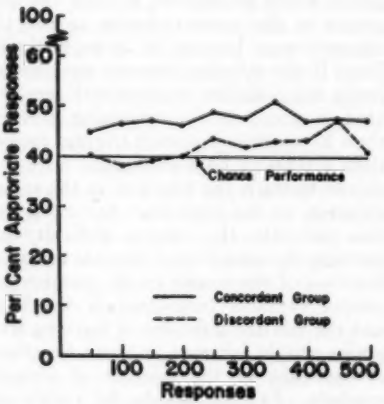


FIG. 4. Mean percentage of appropriate responses with the trigram patterns of Exp. III. Lines on the abscissa represent successive blocks of 50 responses, i.e., responses 1–50, 51–100, etc. Lines on the ordinate refer to the proportion of appropriate responses relative to total responses made.

## EXPERIMENT IV

It was decided that the failure to learn the trigram patterns of Exp. III could have been due to *Ss'* inability to form a workable concept of the (statistical) nature of the learning task. At the end of practice many *Ss* reported that they still were trying to discover and memorize a fixed sequence of events. Consequently, in Exp. IV instructions to *Ss* included a description of the statistical principle by which the trigram patterns were generated. In addition, the number of trials was increased from 500 to 1000.

### *Method*

*Subjects.*—Twelve men and women from elementary psychology courses at The Ohio State University were assigned alternately to two groups of six *Ss* each. There were two men and four women in the Concordant group and four men and two women in the Discordant group.[5] These *Ss* had not served previously in any similar experiment.

*Sequence of correct responses.*—The sequences were the same as those used in Exp. III. For each group the cycle of 50 events was repeated 20 times, instead of 10 times as in Exp. III.

*Procedure.*—Each *S* made a total of 1000 responses in two daily sessions of 500 per day. Rest periods of 5 min. were given at the end of 250 and 750 responses.

A discussion of the principle by which the trigram patterns were generated and an example of a trigram pattern (the example was not a pattern used in the experiment) were added to the instructions of the preceding experiments. Otherwise, the procedure was the same as in Exp. III.

[5] Evaluation of the data of Exp. I by means of chi-square tests indicated a high degree of homogeneity between men and women in their sequential guessing habits.

*Results*

The results are illustrated graphically in Fig. 5. In correspondence with the findings of Exp. II and III, a difference resulted from the use of concordant and discordant trigrams during the early part of practice. This difference was significant beyond the .05 level of confidence ($t = 2.66$; $df = 10$) in favor of the Concordant group for the first 200 responses, but was not significant thereafter.

Examination of Fig. 5 indicates that there was some tendency, particularly for the Discordant group, for the number of correct predictions between trials 500 and 1000 to increase. Examination of individual scores indicated that this tendency can be attributed to an increase in the number of appropriate predictions for three *Ss* in each group. Analyses of variance, however, showed the differences among the means for Trials 1–200, 300–500, and 700–1000 to be significant only between the .10 and .05 levels of confidence for each group; consequently, scores of individual *Ss* were not analyzed further. It can be concluded that learning, even over 1000 trials, was at most very slight.

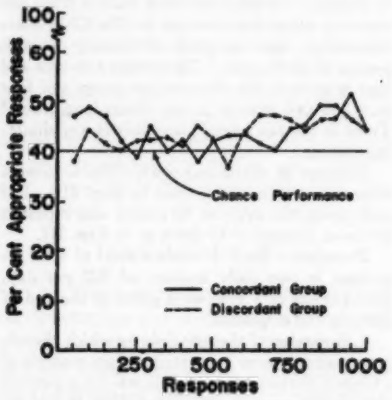


FIG. 5. Mean percentage of appropriate responses with the trigram patterns of Exp. IV. Lines on the abscissa represent successive blocks of 50 responses, i.e., responses 1–50, 51–100, etc. Lines on the ordinate refer to the proportion of appropriate responses relative to total responses made.

## DISCUSSION

Under the conditions of these experiments there was a relatively large and statistically significant amount of learning for the digram sequences, while only a nonsignificant tendency to learn the trigram sequences was present. One possible explanation for the difference between digram and trigram learning is to be found in the fact that there were 50 trigrams but only 10 digrams to be learned. In investigations concerning the relation between the number of items and the number of trials required to learn, positive, negative, and zero accelerations all have been found (4, p. 620). If it is assumed for purposes of illustration that there is a one-to-one relation between number of items in a sequence and the number of learning trials to reach a criterion, then 25, rather than 20, repetitions of each of the 50 trigrams would be required to learn the trigrams to the same criterion as the 10 digrams were learned in 25 repetitions. Even if the relation between number of items and difficulty is negatively accelerated, it would not be surprising if more than 20 repetitions of each trigram (more than a total of 1000 responses) were required to learn the trigrams to the same criterion as the digrams. According to this rationale, the relative difficulty of learning digrams should increase as some function of the square of the number of unique symbols (engrams) in the list, and the relative difficulty of learning trigrams should increase as some function of the cube of the number of unique symbols. In other words, *Ss*' ability to learn higher-order sequential dependencies in this type of learning situation should fall off in an accelerated fashion with the extent of the intersymbol influence.

If it is assumed that ease of learning is inversely related to the total possible

number of events rather than the number actually present in an ergodic sequence, it would also be expected that learning the trigrams would be considerably more difficult than learning the digrams. In a 5-alternative situation there are 125 possible trigrams and only 25 possible digrams, again a ratio of 5 to 1. Thus, if learning is an inverse function, not of the actual number of digrams or trigrams used in a series, but rather of the possible number of alternatives that could have been generated with the number of symbols available, it still would be expected that the present trigram learning situations would require a much greater number of trials than the digram situation in order to reach the same level of performance.

An additional source of increased difficulty lies in the possibility for greater interference as a function of the differential number of events which must be retained (or disregarded) in making predictions as intersymbol influence is lengthened. There may be considerably less intersymbol interference in a situation in which *S* can disregard, in each successive prediction, all but the immediately preceding knowledge-of-results light (digram learning) than when *S* must utilize the first and second knowledge-of-results lights in predicting the third light, then utilize the second and third lights and disregard the first in predicting the fourth light, etc. (trigram learning). In learning trigrams where *S* is set to consider a larger number of previous events, there may also be a tendency for a greater number of previous responses, which result in interference, to be utilized in making the predictions than is the case in digram learning.

Both digram and trigram learning in the particular sequences studied here was further complicated by the fact that at each step in the sequence two next events were always equally probable. Thus, appropriate digram and trigram response sequences were reinforced only 50% of the time, a situation which is known to result in relatively slow learning (5).

Instances can be cited which demonstrate high levels of learning of trigram or higher-order sequences. In such instances, however, there is almost always (*a*) a high degree of overlearning, as in the case of language behavior, and (*b*) unequal engram plus unequal higher-order probabilities (7, 11). Furthermore, and perhaps even more important for learning, in most communication-learning situations individual events in the sequences are grouped either spatially (as in written language) or temporally (as in spoken language) so that *S* responds not to individual events but to sequences of events. The results of the many studies on the influence of grouping in the areas of perception, reading, and verbal learning have made it almost axiomatic that grouping into meaningful units facilitates performance. In the present experiments, on the other hand, the individual events were separated by 4 sec. and *S* was forced to respond between each event. The present results should not be interpreted, therefore, as indicating that trigram or longer dependencies cannot be learned. They indicate only that such learning is extremely difficult when the events in a time series are spaced by as much as 4 sec. and are responded to individually.

One implication of this line of reasoning, which may have important implications for skill learning, is that when sequences (of stimuli or movements) of a statistical nature have to be learned, it may be very important to give knowledge of results in such a way that *S* can observe entire sequences of events.

The influence of sequential guessing habits persisted for a considerable amount of practice, especially in the two situations in which *S*s showed little or no learning. This effect is the more noteworthy in view of the relatively small mean differences in guessing habits found in Exp. I, and the fact that in later experiments the concordant and discordant sequences obtained from the empirical data of Exp. I were employed with different *S*s. These results indicate that sequential guessing habits are an important consideration in the design of many investigations. In fact, it often may be more important to control for

guessing habits than to control for exact frequencies of reinforcement of higher-order sequences.

## Summary

The four experiments reported here are concerned with (*a*) the relative difficulty of learning digram *versus* trigram probabilities, and (*b*) the effect in such situations of sequential guessing habits.

Experiment I was designed to determine sequential guessing habits in a situation where it was possible to make only a chance number of successes. Twenty *S*s predicted which of five events would occur, with knowledge of results provided after each response. By using a restricted random procedure, all possible engrams, digrams, and trigrams appeared equally often. The pattern of correct responses thus had maximum uncertainty.

In Exp. II two groups of 20 *S*s each were given patterns which had equal engram but unequal digram (and trigram) probabilities. For one group the digrams conformed to the digram guessing habits obtained in Exp. I (Concordant group); the digrams for the other group were those used infrequently in Exp. I (Discordant group). A total of 10 digrams was used for both groups: practice continued for a total of 250 responses (25 repetitions of each digram).

Experiments III and IV were similar to Exp. II except that trigram instead of digram patterns were used. All possible engrams and digrams appeared equally often in the pattern of correct responses; an appropriate prediction had to be based on the two preceding events. A total of 50 trigrams was used for both groups in each of these experiments. In Exp. III a total of 500 responses (10 repetitions of each trigram) was given. In Exp. IV the principle by which such patterns are generated was explained, and a total of 1000 responses (20 repetitions of each trigram) was given.

It was found that the digram patterns of Exp. II were learned, but that little or no learning occurred with the trigram patterns of Exp. III and IV. It also was found that the influence of guessing habits persisted for an appreciable number of responses, especially when there was little learning. Possible explanations for the failure to learn the trigram patterns were discussed, as were methodological implications of the finding regarding the influence of guessing sequence habits.

## References

1. Aborn, M., & Rubenstein, N. Information theory and immediate recall. *J. exp. Psychol.*, 1952, **44**, 260–266.
2. Chapanis, A. Random-number guessing behavior. *Amer. Psychologist*, 1953, **8**, 332. (Abstract)
3. Hake, H. W., & Hyman, R. Perception of the statistical structure of a random series of binary symbols. *J. exp. Psychol.*, 1953, **45**, 64–74.
4. Hovland, C. I. Human learning and retention. In S. S. Stevens (Ed.), *Handbook of experimental psychology*. New York: Wiley, 1951. Pp. 613–689.
5. Jenkins, W. O., & Stanley, J. D., Jr. Partial reinforcement: a review and critique. *Psychol. Bull.*, 1950, **47**, 193–234.
6. Miller, G. A., & Frick, F. C. Statistical behavioristics and sequences of responses. *Psychol. Rev.*, 1949, **56**, 311–324.
7. Miller, G. A., & Selfridge, J. A. Verbal context and the recall of meaningful material. *Amer. J. Psychol.*, 1950, **63**, 176–185.
8. Newman, E. B. The pattern of vowels and consonants in various languages. *Amer. J. Psychol.*, 1951, **64**, 369–376.
9. Newman, E. B., & Collier, N. Guessing sequences using two, four, and eight alternatives. *Amer. Psychologist*, 1952, **7**, 272. (Abstract)
10. Senders, V. L., & Sowards, A. Analysis of response sequences in the setting of a psychophysical experiment. *Amer. J. Psychol.*, 1952, **65**, 358–374.
11. Shannon, C. E. Prediction and entropy of printed English. *Bell Syst. Tech. J.*, 1951, **30**, 50–64.

(Received for early publication May 17, 1954)